UNIVERSIDAD AUTÓNOMA DE MADRID

DEPARTAMENTO DE BIOLOGÍA MOLECULAR

In vivo phosphoprteomics reveals kinase activity profiles that predict treatment outcome in triple-negative breast cancer

Ivana Zagorac

Madrid, 2017

DEPARTAMENTO DE BIOLOGÍA MOLECULAR
FACULTAD DE CIENCIAS
UNIVERSIDAD AUTÓNOMA DE MADRID

In vivo phosphoprteomics reveals kinase activity profiles that predict treatment outcome in triple-negative breast cancer

Ivana Zagorac

Licenciatura en Biología Molecular y Fisiología

Director de Thesis:
Miguel Quintela-Fandino, M.D. PhD

This thesis, submitted for the degree of Doctor of Philosophy
at the Universidad Autónoma de Madrid,
has been completed in the Brest Cancer Clinical Research Unit
at the Spanish National Cancer Research Centre (CNIO) under the supervision of. Dr.
Miguel Quintela-Fandino.

Dedicated to my parents,
my sister Sladjana, David and my
friends who always supported me.

ACKNOWLEDGEMENTS

ACKNOWLEADGEMENTS

Looking back, it has been almost eight years since I first entered through the door of CNIO. I remember, I was last year graduate student at that time, when I came to a foreign research centre of excellence like CNIO for the fist time and I was so afraid weather I will be able to fulfil the expectations and manage to live by myself for the first time outside my home country. A year later I remember I was looking for a PhD position and it caught my attention the joining of a young investigator to the CNIO. Thank you Miguel for responding to my email, for encourage me to apply for La Caixa Fellowship and later on giving me the opportunity to develop my scientific carrier in your lab! During these six years I met so many people that I need to mention and give my gratitude for the unconditional support and help.

First of all I have to express my deep gratitude to Paloma. You have been there for me from the first day I started with this adventure. You thought me to look at things with positivity and that with hard work, dedication and patients everything is possible. Tamara, honestly I have not met a person like you because you are unique. No matter how much time will pass I think you would always manage to surprise me. You thought me so many things not only work related but more importantly those that I would carry with me all my life. You would always make things possible even when everyone thought they are impossible. I would always be grateful to you for sharing with me your knowledge. Elena thanks to your patients and willingness to always respond to all my questions I understood how science could be translated to daily clinical practice. You were always ready to address all my doubts and queries, which I think, strengthen my decision to develop my scientific carrier in clinical research field. Juanma, thank you so much for always having time to listen to me and be there to help me understand and overcome any problem I had. I really enjoyed in our scientific discussions and I think this helped me develop the perception of science I have now. Jesus, apart form learning so many things from you, I will be always grateful to you for showing me what I almost forgot, that we always have to value ourselves, and that there will be always more people who would believe in us than other way around. Maria Jose, always somehow in the shadow but always present and ready to help. I admire your modesty and ability for organization and multitasking. Thank you for helping me to come to the end of this journey called PhD.

During this PhD few people continued their scientific carriers outside this lab, but I always felt they following my progress even form the distance, encouraging me to keep going until the target. This opened the possibility for new people to come and leave great professional and personal mark on my carrier. Laura, you have been my lab mate, my running partner, in some way my personal doctor and more importantly an honest and unconditional friend. I admire your innocence and the way you always try to find the good in people. I think you are brilliant and innovative both as medical doctor and scientist. I hope to celebrate with you your thesis because I believe in your talent and brightness. Vero, Espe and Pako I honestly think that without your help and support I would not be able to achieve what I achieved. You all brought the light in my PhD when I needed the most. You gave me strength and supported me whenever I needed. More importantly you have been real friends and I am certain our friendship will last forever. Silvana, thank you for showing me that science could be also fun and that sooner or later every hard work pays back. Sara and Gonzalo thank you for your support during last year of my PhD.

These almost six years brought many more incredible people on my way. I think I could spend ten pages more just to express my gratitude to all of them. Fernando, thank you for showing me the basics of proteomics, for teaching me to write down every step in the experiment, even though I would repeat it more times and a true friend all these years. Nuria, Pilar and Isa I will always be thankful to you, your patients and willingness to respond to every question I had helped me understand complicated world of proteomics. Javier, as I always told you I admire you very much. Thank you for always questioning me and encouraging me to make my own opinions. For being always prepared to give me a hand and showing me that the best way to resolve a problem is to come beck to the basics and ask simple questions. But more importantly thank you for making of me more devoted and passionate scientist.

Gonzalo, thank you for always having time for me, no matter how occupied you were you were always ready to spare few minutes to answer all my questions. Thank you for that.

ACKNOWLEADGEMENTS

Sole thank you for spending with me hours under the microscope trying to se up protocols for immunohistochemistry. Without you help great part of these thesis would not be possible to finish. Thank you also for sharing with me your knowledge.

I have to express my gratitude to all people I did not mentioned but with whom I worked directly or indirectly during these almost six years.

As well, I give my thanks, to people from my PhD committee, Paco Real, David Olmos and Aleix Prat, Thank you for all your support during these years and for the meaningful and productive meetings.

Apart for the people I had the great honour to get to know and work with during these year of my PhD I need to thank other people in my lift who have been there for me no matter the distance or time passed.

I have to thank my best friends Ksenija and Dragana, who always supported me and encouraged me to make decisions I feel best for me. You thought me that real friendship can overcome every obstacle and that time and distance means nothing when you have friend who are your soul mates. David, thank you for being my greatest support during this journey. Thank you for being so understanding when I had to work late hours or during weekends. Thank you for offering me your shoulder any time I needed and telling me that everything will be fine. Thank you for teaching me to have fate in myself and show my strength. Thank you for bringing to my life Ice, the most incredible dog I ever seen, who would always bring the smile on my face no matter how tired or sad I would feel. And lastly thank you for loving me unconditionally. Also I have to thank Consuelo, Jesus and Visi how accepted me and treated me always as part of the family and made me feel that I have always someone to count on.

I have to express my deep gratitude to my incredible parents. You always believed in me through all my life and thought me the fundamental and core values I will bring with me forever. Without your support I would not be where I stand now. I will always look up to you. I hope I will be able to be the same support to you as you have been to me. Moreover special thanks to all my family members. My uncles and aunts, who also always believed in me and supported me. My small cousins, who I hope, some day will

set their journey in the same field as me. Finally, but not less important, I am grateful to my sister Sladjana. You have literarily done this thesis with me. You have been my colleague, my friend and my sister. I will always remember our scientific discussions during lunch break or on our way home. I will always admire you. I believe you are brilliant young scientist, with so much strength and intelligence. Thank you for always keeping my beck and pushing me to reach my goals. I will be always there for you like you were for me, unconditionally.

Lastly, as I started I would like to finish. Thank you Miguel for giving me the opportunity to do my PhD in your lab. It was hard marathon to run, but I managed to finish it! I am more than grateful for every time you very there to say words of encouragement and support. For being tough and demanding and for helping me mature as a scientist and a person.

Још једно желим да се залим мојим родитељима. Иако нисте имали прилике да својим физичким присуством будете уз мене у сваком моменту током ових шест година, осетила сам као да јесте. Ваша подршка је била прекретница у мојој одлуци да наставим школовање у иностранству. Увек ћете бити мој узор и подршка. Слађана не знак како да ти се довољно захвалим за све што си учинила за мене. Хвала ти што никада ниси престајала да верујеш у мене и што си ме увек подсећала да ако нешто заиста желимо у животу, свим срцем, читава беселјана ће се усредсредити да се то и оствари. Хвала ти!

TABLE OF CONTENTS

SUMMARY

Triple-negative breast cancer (TNBC) is am immunohistochemically-defined breast cancer subtype negative for ER, PR and HER2 expression, with dismal prognosis that lacks prognostic and predictive markers. Since all the aberrations that exist and ultimately contribute to the tumor phenotype converge, from a functional point of view, in the final status of the phosphorylation of the proteome in a given moment, we sought to interrogate the phosphoproteome with two aims: 1) establish a taxonomy of TNBC based on measurable markers that predict clinical course; 2) reduce the phosphoproteome that characterizes the bad- from good-prognosis cases to its targetable, driving kinases, in order to define rational therapeutic approaches in TNBC. Therefore we performed mass-spectrometry based quantitative phosphoproteomics from a training set of 34 frozen patient tumor samples paired by tumor size (T), nodal status (N), and grade (G). 13 patients relapsed in less than 3 years and 21 patients were relapse-free for more than 10 years follow up. Moreover in our study we included ten triple-negative cell lines, 3 cell lines that killed recipient mice in less than 100 days and 7 indolent cell lines (no death in more than 400 days). In our study we were able to identify and quantify more than 17000 unique phosphopeptides (pPs) in the patient samples and cell lines. Differential analysis showed that 161 pPs were up-regulated in the relapsed vs non-relapsed group. Furthermore the hyperactivity of 11 kinases accounted for the pPs upregulated in the relapsed group. A mass-spectrometry-to-immunohistochemistry translation step in 113 consecutive TNBC cases, spotted in TMAs, with 14 years follow up revealed that 6/11 kinases (PRKCE, KIT, PNKP, ERK, CDK6 and P70S6K) preserved independent prognostic value. The kinases split the validation set into two patterns: one (0/6 hyperactivated kinases; 29% of the patients) associated to 93.5 % cure rate. The other (≥1 hyperactivated kinase; 37 subpatterns) was associated to a 9.5-fold higher relapse risk. The 15 possible 2-drug combinations targeting these kinases were synergistic in vitro and in vivo. Our results elucidate novel phosposites/kinases implicated in TNBC and provide a ready to use, actionable targets-based clinical classification system for TNBC.

RESUMEN

El cáncer de mama triple negativo (TNBC) es un subtipo de cáncer de mama definido inmunohistoquimicamente como negativo para la expresión de ER, PR y HER2, además carece de marcadores pronósticos y predictivos lo que hace que presente un mal pronóstico en general. Dado que todas las aberraciones genómicas presentes en el tumor contribuyen en el fenotipo tumoral y convergen, desde un punto de vista funcional, en el estado final de la fosforilación del proteoma en un momento dado, se intentó interrogar al fosfoproteoma con dos objetivos: 1) establecer una taxonomía de TNBC basado en marcadores medibles capaces de predecir el curso clínico; 2) simplificar el fosfoproteoma que distingue a pacientes con peor y mejor pronóstico llevado a cabo por las quinasas, con el fin de definir distintos enfoques terapéuticos en el TNBC. Por lo tanto hemos realizado un ánalisis fosfoproteómico cuantitativo, por espectrómetría de masas en un grupo de 34 muestras de pacientes, agrupadas por grado (G), tamaño del tumor (T) y estado de los ganglios (N). 13 de los pacientes recayeron en menos de 3 años y los otros 21 no presentaron recaida en más de 10 años de seguimiento. Además en nuestro estudio se incluyeron diez líneas celurares triple negativas. La inyeccion de 3 de estas líneas celurares en ratones receptores fueron capaces de matarlos en menos de 100 días considerándose líneas agresivas, mientras que las 7 líneas celulares restantes no fueron capaces de matar a los ratones receptores en mas de 400 días y por ello se consideraron no agresivas. En nuestro estudio hemos podido identificar y cuantificar más de 17000 fosfopeptidos únicos, (pPs) presentes en las muestras de pacientes y líneas celulares. El análisis diferencial demostró que 161 de estos pPs idetificados estaban hiperregulados en el grupo de pacientes que recaían comparados con el grupo de pacientes sin recaída. Además, se identificaron 11 kinasas hiperactivas responsables de la hiperregulación de estos pPs, en el grupo de los pacientes que recaian. La validación de los resultados obtenidos mediante espectrometría de masas se llevo a cabo en una neuva seríe de 113 muestras de casos de TNBC (con 14 años de seguimiento) que fueron incluidos en TMAs. Mediante técnicas inmunohistoquímicas observamos que 6 de las 11 kinasas identificadas (PRKCE, KIT, PNKP, ERK, CDK6 y P70S6K) conservaban un valor pronóstico independiente. Estas 6 kinasas son capaces de dividir el grupo de validación en dos patrones: uno asociado a la tasa de curación del 93,5% (0/6 kinasas

hiperactivadas; 29% de los pacientes) y el otro asociado a un mayor riesgo de recaida 9.5-veces (≥1 kinasas hiperactivadas; 37 sub-patrones). El tratamiento con las 15 posibles combinaciones de 2 fármacos dirigidos contra estas kinasas, fueron sinergisticas en tratamientos vitro y en vivo. Nuestros resultados identifican nuevos sitios de fosforilación y kinasas implicadas en el cáncer de mama triple negativo, proporcionando un nuevo sistema de clasificación basado en estos marcadores y que podría ser utilizado en la práctica clínica.

ABBREVIATIONS

TNBC	Triple-negative breast cancer
T	Tumor size
N	Nodal status
G	Tumor grade
pPs	Phosphopeptides
ER	Oestrogen reseptor
PR	Progesteron receptor
HER2	Human epidermal growth factor receptor 2
PARP	Poly - ADP ribose polymerase
CK5/6	Cytokeratin 5/6
EGFR	Epidermal growth factor receptor
CL	Claudin-low subtype
BMI	Lowest body mass index
HR	Hazard ratio
CI	Confidence interval
BLBC	Basal-like breast cancer
PIK3CA	Phosphatidylinositol 3-kinase, catalytic, alfa
USH2A	Ushers syndrome gene
PTEN	Fosfatidilinositol-3,4,5-trisfosfato 3-fosfatasa
RB1	Retinoblastoma
COL6A3	Collagen type VI alpha 3 chain
ATR	Ataxia telangiectasia and Rad3 related
UBR5	Ubiquitin protein ligase E3 component n-recognin 5
BRAF	B-Raf proto-oncogene, serine/threonine kinase
ERBB2	Erb-b2 receptor tyrosine kinase 2
ERBB3	Erb-b2 receptor tyrosine kinase 3
BRCA1	Breast cancer susceptibility gene 1
BRCA2	Breast cancer susceptibility gene 2
RFS	Relapse-free survival
OS	Overall survival
DFS	Disease-free survival
PFS	Progression-free survival
PTMs	Posttranslational modifications
RTKs	Receptor tyrosine kinases
2D-PAGE	2-dimetional-polyacrile amide gel electrophoresis
MS	Mass spectrometry
ICAT	Isotope-coded affinity tag
SILAC	Stable isotope labelling by amino acids in cell culture
iTRAQ	Isobaric tags for relative and absolute quantification
LC	Liquid chromatography
Ser (S)	Serin
Thr(T)	Threonin
Tyr(Y)	Tyrosine
DMEM	Dulbecco's Modified Eagle Medium
EDTA	Ethylenediaminetetraacetic acid
FBS	Fetal bovine serum

FFPE	Formalin-Fixed, Paraffin-Embedded
HRP	Horseradish peroxidase
HNU mice	Athymic Nude-Foxn1nu
PBS	Phosphate buffered saline
PEN	Penicillin
RIPA	Radioimmunoprecipitation assay
RNA	Ribonucleic acid
RPMI	Roswell Park Memorial Institute medium
STREP	Streptomycin
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
CDK6	Cyclin-dependent kinase 6
PRKCE	Protein kinase C epsilon
AKT	RAC-alpha serine/threonine-protein
PNKP	Bifunctional polynucleotide phosphatase/kinase
PIM1	Proto-oncogene Serine/threonine-protein kinase
CLK1	Dual specificity protein kinase CLK1
DAPK3	Death-associated protein kinase 3
CHK1	Checkpoint kinase 1
S6	S6 ribosomal protein
ERK1/2	Extracellular signal-regulated kinase
CDK1	Cyclin-dependent kinase 1
P70S6K	Phosphoprotein 70 ribosomal protein S6 kinase

INTRODUCTION

1. BREAST CANCER

Even though being the subject of intensive investigation, breast cancer remains the most common cancer among woman, with greater than 1,300,000 cases and 450,000 deaths each year worldwide[1]. In the last decades intensive research has led to better understanding of breast cancer origin, pathology and treatment options. These efforts have resulted in substantial improvement in the survival rate of breast cancer patients considering that the rate has increased up to 77% by the mid-nineties (Figure 1). Unfortunately, the current long-term prognosis is not favourable and the death rates are still very high[2]. This highlights that more sensitive methods to detect breast cancer are needed and moreover that is extremely important to understand very early events of breast cancer biology in order to improve treatment modalities.

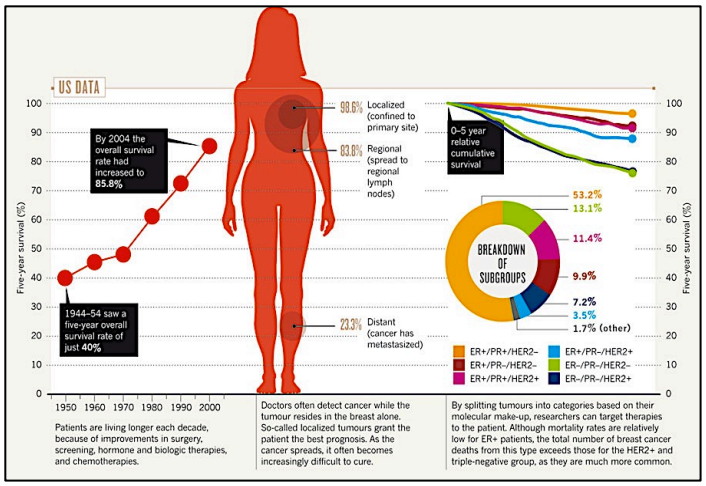


Figure I1. Overall survival rate of women diagnosed with breast cancer (adapted from Maxmen et al. [3])

1.1. Risk factors for Breast Cancer

Although several risk factors for breast cancer have been documented, it is very difficult to identify specific risk factors for most women diagnosed[4]. Risk factors for this disease can be grouped in three categories: hereditary, hormonal and reproductive and environmental factors[5].

A familiar history of breast cancer increases the risk by a factor of two or three. In particular the BRCA1 gene has been implicated in the pathogenesis of hereditary breast

cancer whereby women with mutations in either BRCA1 or BRCA2 have a 60% to 85% lifetime risk for developing breast cancer[6]. Mutations in p53 account for 1% of early onset breast cancer[7], and thus account for a small portion of the total breast cancer burden. Beck in 1994, Guinee at al. were the first to suggest a link between breast cancer and hormonal and reproductive factors. These authors showed the association between the incidence of breast cancer and the age of menarche, first pregnancy and menopause. In particular, those woman who become pregnant after the age of 35, have a two to threefold higher risk of breast cancer than those women whose pregnancy occurred before age 25[8]. Moreover environmental factors including smoking, dietary factors and different chemicals have also been found to influence the risk of breast cancer[9].

1.2. Classification of Breast Cancer

Breast cancer is considered to be a highly heterogeneous disease. Indeed, different types of this cancer exhibit different histopathological and biological features, clinical outcome and response to systemic interventions. Thus, due to this high degree of heterogeneity, this disease cannot be viewed as a single clinico-pathological entity, but rather must be dissected into a number of more homogeneous groups. For classification of any disease, a suitable one has to be easily applicable, widely reproducible and most importantly clinically useful. When it comes to breast cancer, unfortunately, despite all the efforts in past and recent years, the 'perfect' classification is still lacking.

1.2.1. Histological types of breast cancer

The histopathological classification of breast tumors is based on the diversity of its morphological features. Broadly, breast cancer can be categorized into in situ carcinoma (if the tumor is limited to the epithelial component) and invasive (infiltrating) carcinoma (if the tumor has invaded the stroma). Moreover, histological appearance depends on the organ compartment in which the tumor has originated, such as inner lining epithelium of the ducts (ductal carcinoma) or the lobules that supply the ducts with milk (lobular carcinoma) [10].

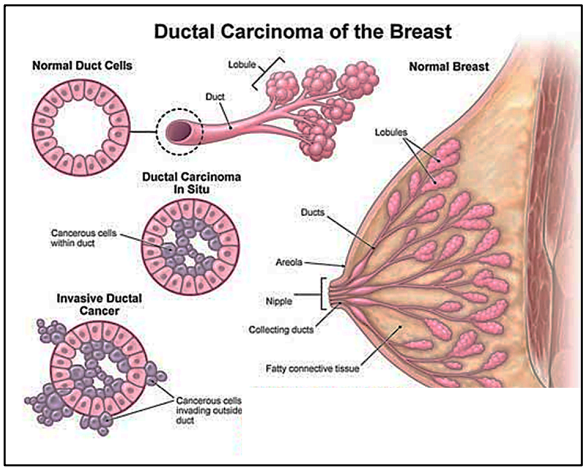


Figure I2. Ductal carcinoma of the breast (adapted from http://ungbuouvietnam.com/ung-thu-bieu-mo-tuyen-vu-xam-lan/)

Adenocarcinomas are found in more than 95% of the patients[11] and invasive ductal carcinoma (IDC) is the most common form of invasive breast cancer. It accounts for 55% of breast cancer incidence upon diagnosis[12].. While this classification scheme has been a valuable tool for several decades, it relies solely on histology without utilizing newer molecular markers that have a proven prognostic significance.

1.2.2. Molecular types of breast cancer

In the last decade, many efforts have been made to substitute the histopathological classification with molecular parameters that can both provide better prediction of tumor behaviour and improve therapeutic strategies. In line with this, Perou and colleagues were able to identify molecular defined classes of breast cancer using hierarchical clustering analysis and similarities in gene expression profiles[13] . The subtypes of breast cancers identified by their gene signature are:

Luminal A - accounts for 50% of invasive breast cancers. It is oestrogen and progesterone (ER/PR) positive or HER2 negative. It has a good prognosis and is typically of low pathological grade and low proliferation rates[14].

Luminal B - comprises 20% of invasive breast cancers. The oestrogen/progesterone expression (ER/PR) is positive, while human epidermal growth factor receptor 2 (HER2) expression is variable (positive or negative). The proliferation index expressed by Ki-67

staining and histological grades are higher than luminal A. The response to endocrine therapy and chemotherapy is variable, and its prognosis is poorer than luminal A[14].

Her2 overexpression - accounts for 15% of all invasive breast cancers. The ER/PR is usually negative, while HER2 is strong positive. It's characterized by high Ki-67 expression and common mutation of TP53. This group of breast cancer implies poor prognosis and shows the highest sensitivity to trastuzumab (herceptin) therapy.

Basal like – this class is so named due to its pattern of expression that is similar to basal epithelial cells and normal myoepithelial cells of mammary tissue. Typically is cytokeratin 5/6 (CK5/6) and/or EGFR positive, ER/ PR negative, and HER2 negative (triple negative) with a high expression of Ki-67 and common TP53 mutations. In terms of treatment options, this type of breast cancer do not respond to endocrine therapy or trastuzumab but shows some sensitivity to platinum-based chemotherapy and poly - ADP ribose polymerase (PARP) inhibitors. In general has poor prognosis.

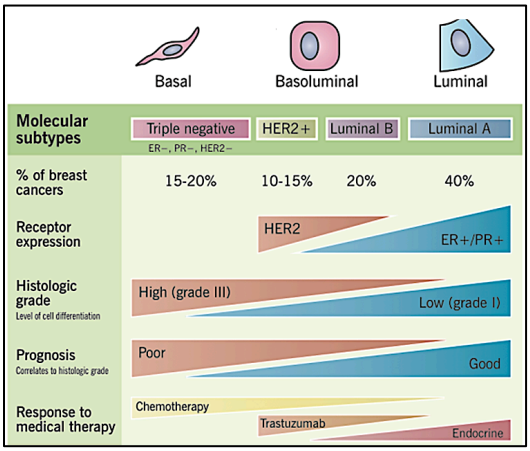


Figure I3. Molecular types of breast cancer (adapted from http://www.pathophys.org/breast-cancer/)

It should be noted that this classification based on molecular and immunohistochemical parameters do not overlap completely, since some basal-like breast cancers (according to the molecular classification) will not show the expected triple-negative (ER, PR and HER2 negative) immunophenotype and vice versa, not all triple-negative breast cancer will be classified as basal-like by gene expression profiling.

In 2007, a new intrinsic subtype was described, the claudin-low subtype (CL), through the combined analysis of murine mammary carcinoma models and human breast cancers[15]. This subtype represented 6% of the breast cancer samples analyzed. CL tumors lack tight junction proteins including claudin 3 and E-cadherin[15] and are characterized by low to absent expression of luminal differentiation markers, high enrichment for epithelial-to-mesenchymal transition markers, immune response genes and cancer stem cell-like features.

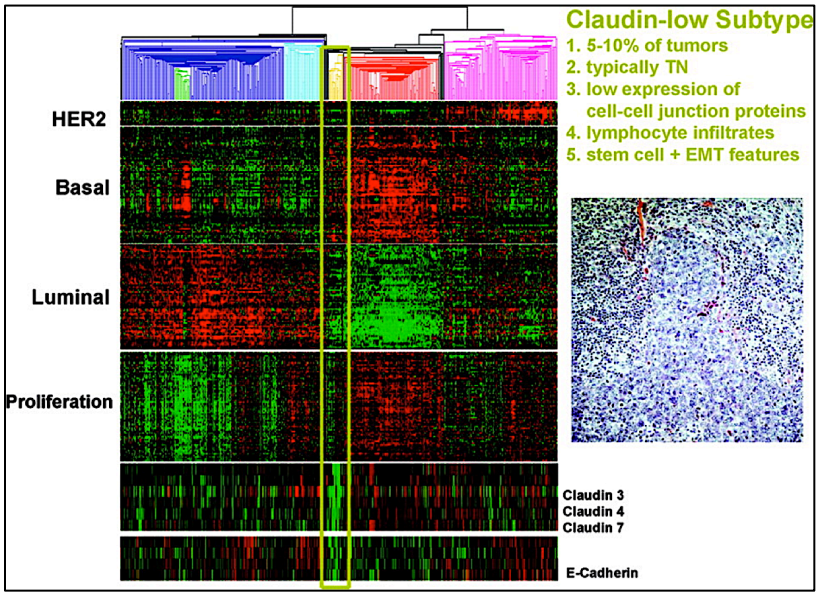


Figure I4. Cluster diagram of intrinsic subtypes highlighting the claudin-low subtype. Abbreviations: EMT, epithelial–mesenchymal transition; HER-2, human epidermal growth factor receptor; TN, triple negative. (Adapted from Perou et al[16])

Clinically, the majority of claudin-low tumors are poor prognosis oestrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and epidermal growth factor receptor 2 (HER2)-negative (triple negative) invasive ductal carcinomas with a high frequency of metaplastic and medullary differentiation. They also have a response rate to standard preoperative chemotherapy that is intermediate between that of basal-like and luminal tumors[15,17,18]. Today, with less than 90 samples characterized, the CL subtype is the least characterized subtype in the literature.

2. TRIPLE-NEGATIVE BREAST CANCER

Triple negative breast cancer (TNBC) has recently been recognized as an important subgroup of breast cancer with a distinct outcome and therapeutic approach when compared to other subgroups of breast cancer. This subtype is characterized by the absence of oestrogen and progesterone receptor expression and no overexpression of human epidermal growth factor receptor 2 (HER2). In different series and patient populations TNBC may range between 6–28% of breast cancers[19–21], but even higher incidence rates are reported for some ethnical groups such as African Americans and for younger patients[22]. Although it accounts for a relatively small minority of breast cancer cases (10–17 %)[23], TNBC is responsible for a disproportionate number of breast cancer deaths. Moreover, there have been fewer advances in the treatment of TNBC than of other subtypes. For these reasons, new research initiatives for TNBC are critical[24].

2.1. Risk factors for Triple Negative Breast Cancer

Two studies have shown that reproductive factors — specifically, pregnancy and multiple childbirth, obesity and lack of physical activity increase the risk for TNBC. In the first study, the researchers found that obese women had a 35% higher risk for TNBC than women with a lowest body mass index (BMI); in the second study, women who had never given birth had a 40% lower risk for the disease than those who had experienced a full-term pregnancy. Surprisingly, the number of births affected the risk for TNBC; women who had given birth to 3 or more children were at higher risk than women who had given birth to 1 child (hazard ratio [HR], 1.46; 95% confidence interval [CI], 0.82 to 2.63)[25].

2.2. TNBC versus BLBC

Triple negative breast cancers is described as a type of breast cancer that lacks expression of oestrogen receptor (ER), progesterone receptor (PR) and HER2. Moreover they are characterized by its similarity to basal-like cancers including the fact that they more frequently affect younger patients (<50 years)[19,26,27], are more prevalent

in African-American women[28], often present as interval cancers and are significantly more aggressive than tumors of other molecular subtypes[19,29]. This aggressiveness is best exemplified by the fact that the pick of recurrence is between the first and the third year and the majority of deaths occur in first 5 years following treatment[21,30]. Patients with triple negative cancers[21], similar to those with basal-like cancers have significantly shorter survival following the first metastatic event when compared with those with non-triple-negative/non-basal like controls.

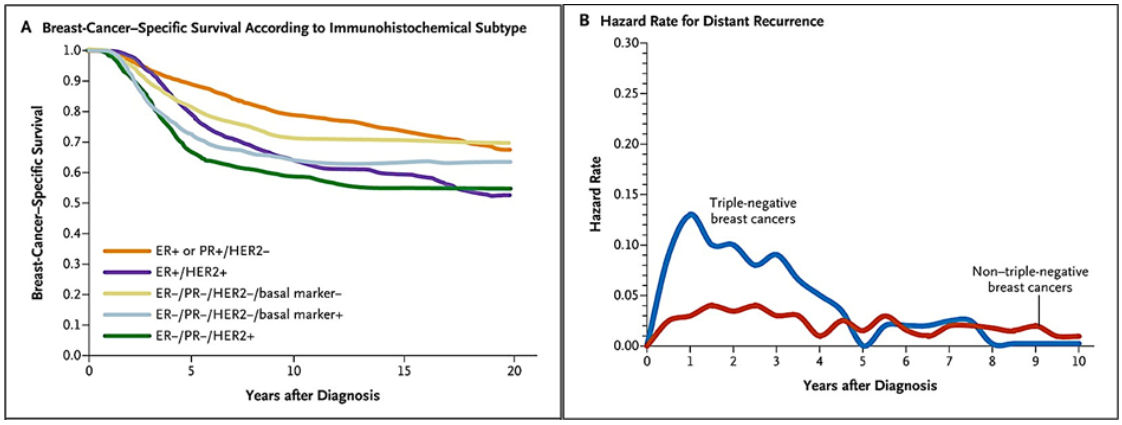


Figure I5. Survival after a Diagnosis of Breast Cancer. Panel A shows the survival rate in a series of 3744 patients according to immunohistochemical subtype. Panel B shows the hazard rates for distant recurrence of triple-negative breast cancer and non-triple negative breast cancer (adapted from Foulkes et al.[31])

Several studies have shown that majority of TNBC cancers are of basal phenotype[20,32,33] and vice versa that majority of tumors expressing 'basal' markers are triple-negative[20,34–36]. However, not all basal-like cancers determined by gene expression profiling lack ER, PR and HER2[34,36–39] and conversely not all triple-negative cancers show a basal-like phenotype by expression array analysis. For example, Bertucci *et al*[36] showed that only 71% of triple-negative cancers were of basal-like gene expression profile. On the other hand only 77% of molecular basal-like tumors were triple-negative[40,41]. Further, there is a significant number of triple-negative cancers that do not express basal markers and are classified as other breast cancer subgroups, such as normal breast-like[41–44] or claudin-low[15,45]. These findings suggest that when analyzed by gene expression profiling TNBC do not form homogeneous group. In contrast, the basal-like subtype does form a homogeneous group of tumors with a

similar gene expression profiling related to prognosis and therapy response[36,46]. Thus, the overall poor prognosis of TNBC may be a result of this basal-like subgroup, and triple negativity may be seen more as a symptom than as a separate entity of breast cancer. In line with this, triple-negative phenotype may not be an ideal surrogate marker for basal-like breast cancer[15,36].

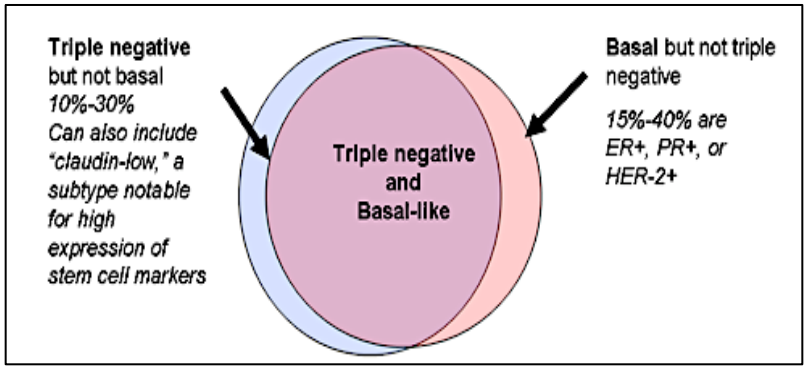


Figure I6. Overlap between triple-negative and basal-like breast cancer (adopted from http://www.medscape.org/viewarticle/587289)

2.3. Molecular characteristics of triple-negative breast cancer

Genomic profiles of TNBC have recently been defined using whole exome sequencing which revealed that TNBC patients harbor germ line mutations in tumor relevant genes[47], such as PALB2, but also high numbers of somatic mutations[48,49]. These studies together with others[50,51], have discovered that one of the most frequently mutated genes is p53 with 62% of basal TNBC (determined by PAM50[41] analysis on RNASeq expression profiles) and 43% of non-basal TNBC cases harboring a validated somatic mutation. The same study also found frequent mutations in *PIK3CA* at 10.2%, *USH2A* (Ushers syndrome gene, implicated in actin cytoskeletal functions) at 9.2%, *PTEN* and *RB1* at 7.7% (5/65) and a further 8 genes (including *ATR, UBR5* (*EDD1*), *COL6A3*) at 6.2% (4/65) of cases in the cohort. Approximately 20% of cases contained examples of potentially "clinically actionable" somatic aberrations, including *BRAF* V600E, high-level *EGFR* amplifications and *ERBB2/ERBB3* mutations. Moreover, both TNBC and basal-like breast cancer show considerable overlap with BRCA1-mutated tumours[52]. BRCA1 is an important tumor suppressor gene that plays a vital role in DNA

repair[22]. Thus the lack of BRCA1 results in DNA repair by more error-prone mechanisms, resulting in genomic instability. Clustering analysis of microarray RNA expression data have shown that familial-BRCA1 mutant tumors strongly segregate with triple-negative sporadic tumors, suggesting similar carcinogenic pathways or cause of these two subtypes[53]. BRCA1 deficiency (as assessed by mutation, methylation, and BRCA1-like comparative genomic hybridization (CGH) profile analysis) is observed in ~50% of patients with TNBC.

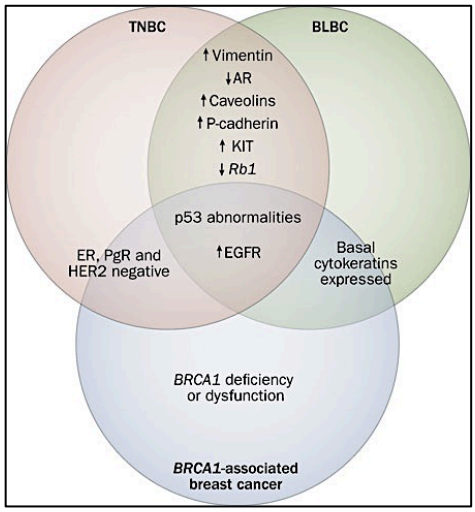


Figure I7. Shared features of triple-negative, basal-like and BRCA1-associated breast cancers (adopted from Carey et al.[54].

BRCA1-associated cancers show a further similarity with triple-negative cancer, in that p53 is commonly mutated, although the spectrum on mutations is distinct from that occurring in triple-negative breast cancer.

There is increasing evidence that the DNA-repair defects characteristic of BRCA-related cancer, especially defective homologous recombination, confer sensitivity to certain systemic agents[55,56], which could be relevant to the treatment of triple-negative breast cancer. In vitro data have shown that inhibition of ADP-ribose-polymerase (PARP)-1 leads to highly selective apoptosis of BRCA1 null cells[57]. At this regard, in December 2014, the FDA approved olaparib (PARP inhibitor) as monotherapy for the treatment of patients with germline BRCA-mutated advanced ovarian cancer who have been treated with three or more prior lines of chemotherapy. In addition there are active phase III studies for olaparib and other PARP inhibitors in metastatic breast cancer for patients

with BRCA1/2 mutations. Nevertheless, it remains unclear whether a positive test for BRCA1/2 mutations can lead to a reliable determination about which patients will benefit from PARP inhibitor therapy.

2.4. Clinical features of TNBC

TNBC is characterized as an aggressive disease with poor clinical outcomes, including short durations of relapse-free survival (RFS) and overall survival (OS)[53]. Data show that a high proportion of patients with TNBC (~34%) experience distant recurrence, with a mean of 2.6 years to distant recurrence, whereas in other breast cancer subtypes distant recurrence rate is of 20%, with a mean of 5 years to distant recurency[21]. The 5-year OS rate is also inferior for patients with TNBC compared with the rate of 5-year OS observed for other types of breast cancer. In TNBC, the risk of recurrence peaks within the first 3 years after diagnosis and then decreases[21,58]. Patients with TNBC are most likely to experience recurrence in the visceral organs. In line with this, study done by Liedtke and colleagues found that 74% of patients with TNBC developed recurrences in visceral organs; but only 13% of patients developed recurrences in the soft tissue and in bone[58].

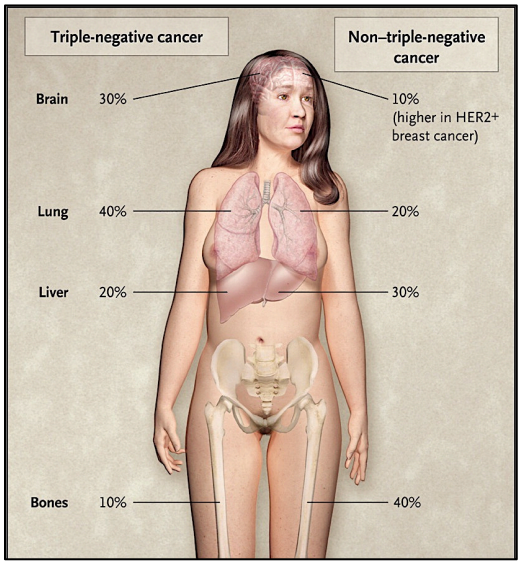


Figure I8. Sites of First Distant Recurrence in Cases of Metastatic Triple-Negative Breast Cancer as Compared with Non–Triple-Negative Breast Cancer. The percentages shown are approximate percentages of women with a first distant recurrence among women in whom metastases develop [31]

Furthermore, TNBC tend to be grade III and large (>2.0 cm) at diagnosis[21]. They have been associated with a higher rate of lymph node positivity than other breast cancer subtypes, but there is not apparent correlation between tumor size and nodal status.

2.5. Treatment of TNBC

TNBC is a complex disease because no dominant oncogenic factor has been identified as a driver of tumor cell proliferation. It is well established that TNBC is an aggressive group of breast cancer subtypes despite having a good initial response to chemotherapy. Moreover, it was found that patients with residual TNBC disease post neo-adjuvant chemotherapy have a worse prognosis than those with non-TNBC[29]. Importantly, there is no preferred standard chemotherapy for these patients and typically tumour size, lymph node status, grade and overall performance status will determine the regimens used. For women presenting with triple negative breast cancer (TNBC), it is established that in the absence of ER, PR and HER-2, endocrine therapies such as tamoxifen and aromatase inhibitors and as well as HER-2 directed therapies such as trastazumab and lapatinib are not efficacious. However, triple negative breast cancers seem to be particularly chemo-sensitive to anthracyclines and taxanes that are part of the standard therapy used for high-risk patients, for example those patients with node positive disease. Unfortunately, in spite of this, there is an overall poorer survival[59].

Neoadjuvant chemotherapy can be used to improve subsequent surgical intervention and/or for testing novel therapies or combinations. Historically, TNBC patients responded well in the neoadjuvant setting, with rates of pathological complete response (pCR) commonly higher than those observed for other breast tumour types[60].

Published trials have also shown that TNBC patients benefit from adjuvant chemotherapy. In the prospective BCIRG 001 study, a chemotherapy regimen comprising combination of docetaxel/doxorubicin/cyclophosphamide (TAC) was compared with fluorouracil/doxorubicin-cyclophosphamide (FAC) in patients with lymph node–positive breast cancer[61]. This study showed that TNBC patients had a better

response to the TAC regimen than to the FAC regimen; 3-year DFS rates were 73.5% versus 60% (HR, 0.50; 95% CI, 0.29-1.00; P = .051). Thus, the TAC regimen may be beneficial as a treatment for TNBC patients, but this regimen is quite toxic and is not tolerated by all patients. Also, studies have not yet explored how these results compare with other anthracycline/taxane sequential chemotherapy regimens.

In metastatic disease, Miller and colleagues compared the efficacy and safety of paclitaxel alone versus paclitaxel plus bevacizumab (Avastin), a monoclonal antibody targeting vascular endothelial growth factor (VEGF)-A as an initial treatment of HER2-negative, metastatic breast cancer. Compared with paclitaxel alone, the bevacizumab/paclitaxel combination demonstrated significantly prolonged PFS (5.9 mo vs 11.8 mo, respectively; P <.001) and an increased objective response rate (ORR; 21.2% vs 36.9%, respectively; P <.001). Patients whose tumors were ER- and PR-negative had significantly improved PFS with the combination regimen than with paclitaxel monotherapy (8.8 mo vs 4.6 mo, respectively; HR, 0.53; 95% CI, 0.40-0.70). Other studies have compared outcomes using different chemotherapy agents in patients with metastatic TNBC. A phase III trial by Thomas and colleagues randomly assigned 752 patients with advanced or metastatic breast cancer who were pretreated or resistant to anthracycline- and taxane-based regimens to receive ixabepilone plus capecitabine (Xeloda) or capecitabine alone[62]. Overall, the ixabepilone/capecitabine combination was more effective than capecitabine alone in prolonging median PFS (5.8 mo vs 4.2 mo, respectively; P = .0003) and was associated with a 25% reduction in the estimated risk of disease progression (HR, 0.75; 95% CI, 0.64-0.88; P = .0003). These results further emphasize the overall sensitivity of TNBC to standard chemotherapy regimens.

In order to improve treatment options for TNBC, researchers have been extensively studied different molecular targets. These include molecular entities, such as expression of HER1 and c-Kit, mutation/perturbation of p53, activation of BRCA1-associated pathways, namely PARP1; and activation of protein kinase components of the mitogen-activated protein kinase and protein kinase B (Akt) pathways[35,53,63,64]. In addition, GRB7, a calmodulin-binding protein that binds phosphorylated tyrosine

residues (ie, EGFR, HER2) and the small heat shock protein α-basic-crystallin (αB-crystallin) have been associated with poor outcome among patients diagnosed with triple-negative and basal-like breast tumors[65,66]. Strategies targeting the EGFR and Src pathways, PARP1 inhibitors, and antiangiogenic agents are currently in clinical trials and are summarized in Table 2.

Therapeutic Strategy or Target	Status of Development
Anthracycline-/taxane-based chemotherapy	Proven efficacy, phase II/III clinical trials (134-131)
Platinum agents	Active agents, phase II clinical trials (138-140)
EGFR inhibition (Cetuximab)	Modest activity, phase II clinical trials(138-140)
Antiangiogenesis (Bevacizumab)	Efficacy in subset analysis, phase III trials (141,142)
PARP1 inhibition (BSI-201, AG014699, AZD2881)	Safety illustrated, efficacy results anticipated, phase I/II trials (140,143,144)
Src inhibition (Dasatinib)	Modest activity, phase II trials (145)
MEK inhibition	Activity in preclinical studies (146)

Table I1. Stratages targeting the EGFR and Src pathways, PARP1 inhibitors and angiogenic agents

3. PROTEOMICS IN CANCER

Cancers are highly complex and are often driven by gene mutations that result in constitutively active signalling pathways that controle cell proliferation and growth[67]. Even though, many genome wide studies illustrated the genetic basis of cancer, they however, do not allow an accurate prediction of what is actually occurring at the protein level within a given cell at any given time. Indeed, studies focusing on detecting the differential expression of mRNA expression have been extremely informative, but they do not necessarily correlate with the protein function. Macromolecules, in general, and proteins, in particular, are highly dynamic molecules. In the other words, any protein may exist in multiple forms that can differ within a particular cell or between different cells. These differences can be the result

of translational, posttranslational, regulatory or degradation processes that can affect protein structure, localization, function and turnover[68].

The deregulation of mechanisms controlling posttranslational modifications (PTMs) can have severe consequences. For instance, mutations in receptor tyrosine kinases (RTKs) can induce constitutive activation of the processes they regulate. In fact, many tyrosine kinases are oncogenes, and malfunctioning survival or differentiation mechanisms will ultimately lead to malignant transformation. Therefore, identification of the kinases that are active in certain tumours is indispensable, as this activation is not always directly caused by genetic mutation[67]. In-depth studying of proteomics profiles of various biospecimens obtained from cancer patients are expected to increase our understanding of tumor pathogenesis and improve the identification of novel targets for cancer therapy. In addition, an essential goal for applying proteomics to study cancers is to adapt its high-throughput tools for regular use in clinical laboratories for the purpose of diagnostic and prognostic categorization of cancers, as well as in assessing various cancer therapeutic regimens[68].

3.1. Mass spectrometry-based proteomics

Global protein analysis poses a tough analytical challenge, in part owing to the highly diverse physicochemical properties of amino acids, which are the building blocks of proteins. Furthermore, compared to the genome, the proteome is complemented by alternative splicing and diverse protein modifications and degradation, and this complexity is further amplified by the interconnectivity of proteins into complexes and signalling networks that are highly divergent in time and space. Quantitative proteomics has been used to identify and quantify proteins in complex biological samples[69].

The classical method for quantitative analysis of protein mixtures is by protein separation and comparison by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), followed by MS or MS/MS[70,71]. However, 2D-PAGE technique suffers from its inability to analyse hydrophobic, and very high or low molecular weight proteins[69]. Moreover, this technique remains a labour-intensive approach, requiring several different experiments for high-throughput studies.

To address this, non-gel-based quantitative proteomics methods have been developed to widen the protein dynamic range and profile[69,72]. These approaches utilise isotope-labelled compounds that are identical to the properties of their natural compounds except in their mass that allows for their identification in mass spectrometry. Stable labelling approaches include isotope-coded affinity tag (ICAT), stable isotope labelling by amino acids in cell culture (SILAC), isobaric tags for relative and absolute quantification (iTRAQ), 15N/14N metabolic labelling and other chemical labelling[73,74]. These approaches are coupled to liquid chromatography (LC) for separation before MS or MS/MS identification. Some limitations of the above approaches include increased sample preparation time, more complex methodology and higher costs attributed to labelling reagents[69]. Furthermore, simultaneous quantification using labelling methods is only possible between few samples[69,72]. Label-free quantification is an alternative to often costly labelling strategies in complex samples like cancer tissues. The use of this approach has increased enormously in the past 10 years and has shown potential for identification and quantification of differentially expressed proteins between normal and tumour samples. Therefore using label-free method in complex samples is a key tool in biomarker discovery.

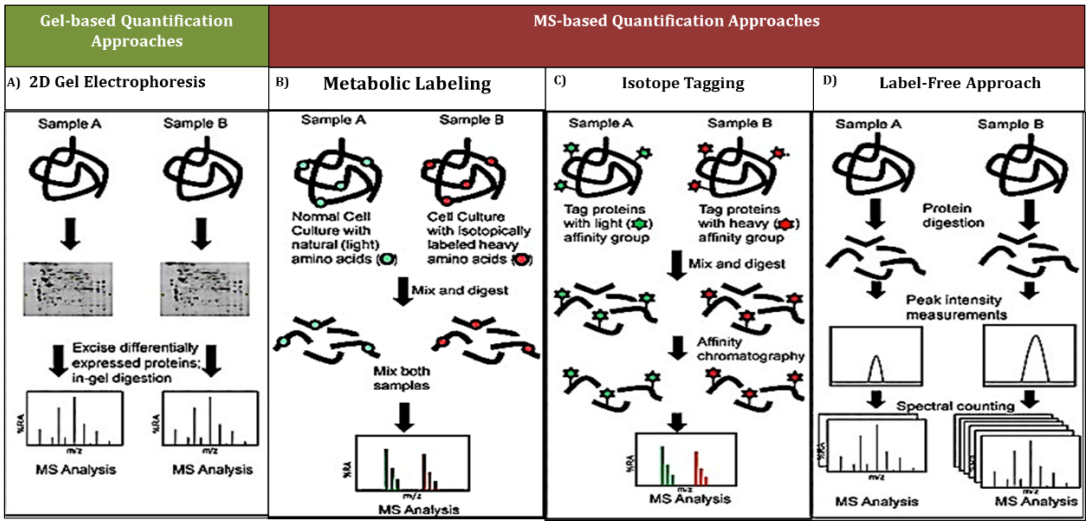


Figure I9. General depiction of various strategies in differential protein expression analysis: **A** is one of the gel-based approaches, where 2D gel electrophoresis is performed to identify differential protein expression on the same gel with fluorescently labeled proteins. The gel spots of interest are picked and in-gel digested for MS analysis for protein identification. **B** through **D** are mass spectrometry (MS)-based quantification approaches classified as **B**, metabolic; **C**, chemical tagging, and **D**, label-free approaches, based on the method of quantification

3.1.1. Label-free protein quantification

Label-free quantification overcomes the expensive and extensive workflows required in the labeling techniques. It is based on the correlation between peptide MS peak intensity and the abundance of the protein in a given sample. Regardless of which label-free quantitative proteomics method is used, they all include the following fundamental steps: (i) sample preparation including protein extraction, reduction, alkylation, and digestion; (ii) sample separation by liquid chromatography (LC or LC/LC) and analysis by MS/MS; (iii) data analysis including peptide/protein identification, quantification, and statistical analysis. In protein-labeling approaches, different protein samples are combined together once labeling is finished and the pooled mixtures are then taken through the sample preparation step before being analyzed by a single LC-MS/MS or LC/LC-MS/MS experiment (Figure 11(a)). In contrast, with label-free quantitative methods, each sample is separately prepared, then subjected to individual LC-MS/MS or LC/LC-MS/MS runs (Figure 11(b)). Protein quantification is generally based on two categories of measurements. In the first are the measurements of ion intensity changes such as peptide peak areas or peak heights in chromatography. The second is based on the spectral counting of identified proteins after MS/MS analysis. Peptide peak intensity or spectral count is measured for individual LC-MS/MS or LC/LC-MS/MS runs and changes in protein abundance are calculated via a direct comparison between different analyses[72].

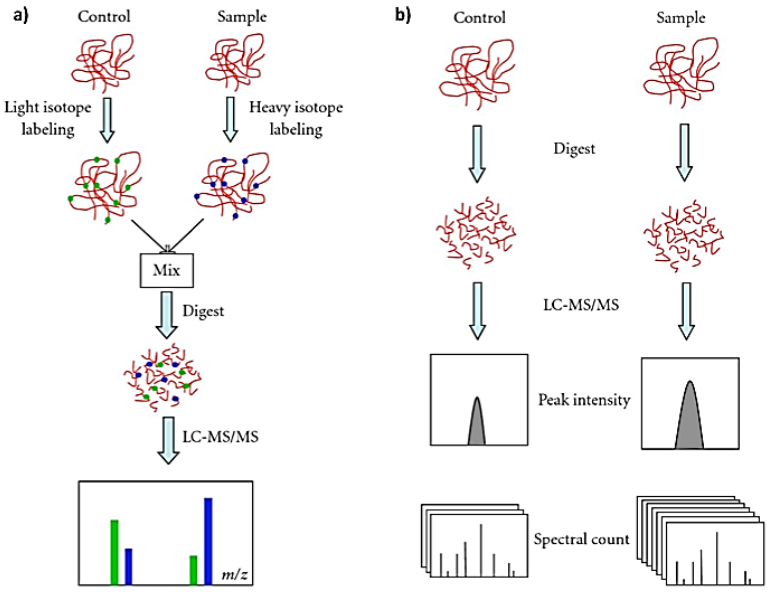


Figure I10. General approaches of quantitative proteomics. (a) Shotgun isotope labelling method. After labelling by light and heavy stable isotope, the control and sample are combined and analysed by LC-MS/MS. The quantification is calculated based on the intensity ratio of isotope-labelled peptide pairs. (b) Label-free quantitative proteomics. Control and sample are subject to individual LC-MS/MS analysis. Quantification is based on the comparison of peak intensity of the same peptide or the spectral count of the same protein (Adopted from Zhu et al.[72])

For the relative quantification by peak intensity of LC-MS, an ion with a particular m/z is detected and recorded with a particular intensity, at a particular time. It has been observed that signal intensity from electrospray ionization (ESI) correlates with ion concentration[75] . The label-free quantification of peptide/protein via peak intensity in LC-MS was first studied by loading 10 fmol–100 pmol of myoglobin digests to nano-LC and analyzing by LC/MS/MS[76]. When the chromatographic peak areas of the identified peptides were extracted and calculated, the peak areas were found to increase with increased concentration of injected peptides. After the peak areas of all identified myoglobin peptides were combined and plotted against the protein amount, the peak area was found to correlate linearly to the concentration of protein. The strong correlation between chromatographic peak areas and the peptide/protein concentration remained when myoglobin was spiked into a complex mixture (human serum) and its digests were detected by LC-MS/MS. The results of quantitative profiling were further improved by normalizing the calculated peak areas[76,77].

Automatic comparison of peak intensity from multiple LC-MS datasets is well suited for clinical biomarker discovery, which normally requires high sample throughput. The following studies were all performed using this label-free quantitative approach. The comparison of control and radiated human colon cancer cells proved the reproducibility of this label-free approach[78]; the serum proteomic profiling of familial adenomatous polyposis patients revealed multiple novel celecoxib-modulated proteins[79]; proteins significantly associated with metastasis were identified by analyzing paraffin-embedded archival melanomas[80]; the analysis of 55 clinical serum samples from schizophrenia patients and healthy volunteers identified hundreds of differentially expressed serum proteins[81]; diagnostic markers and protein signatures were recognized from the serum of Gaucher patients[82] and the cerebrospinal fluid of schizophrenia patients[83].

In the spectral counting approach, relative protein quantification is achieved by comparing the number of identified MS/MS spectra from the same protein in each of the multiple LC-MS/MS or LC/LC-MS/MS datasets. This is possible because an increase in protein abundance typically results in an increase in the number of its proteolytic peptides, and vice versa. This increased number of (tryptic) digests then usually results in an increase in protein sequence coverage, the number of identified unique peptides, and the number of identified total MS/MS spectra (spectral count) for each protein[84]. Liu et al. studied the correlation between relative protein abundance and sequence coverage, peptide number, and spectral count. It was demonstrated that among all the factors of identification, only spectral count showed strong linear correlation with relative protein abundance with a dynamic range over 2 orders of magnitude[85].

Relative quantification by spectral count has been widely applied in different biological complex, including analysis of urine sample from healthy donors and patients with acute inflammation[86], finding biomarkers in human saliva proteome in type-2 diabetes[87], comparison of protein expression in yeast and mammalian cells under different culture conditions[88–90], distinguishing lung cancer from normal[91], screening of phosphotyrosine-binding proteins in mammalian cells[92], and identifying differential plasma membrane proteins from terminally differentiated mouse cell lines[93].

3.2. Phosphoproteomics

Posttranslational modifications (PTMs) are key regulators of protein activity and involve the reversible covalent modification of proteins by small chemical groups, lipids or even small proteins. In addition, proteins can be cleaved by proteases, and the chemical nature of amino acids can be modified. Taking into account the number of expressed protein-coding human genes (~11,000), the array of PTMs available (more than 200), the number of potentially modified residues, the dynamic nature and the often low stoichiometry of these modifications, one realizes the magnitude of the analytical challenge to identify and to localize the sites of these modifications[67]. Probably owing to its ubiquitous role in almost any biological process, phosphorylation is still the best-characterized PTM. Protein phosphorylation is of central importance in signalling systems and employed by cells to transiently alter protein properties such as the activity of enzymes, their interactions with other proteins, their localization and conformations, or to target them for destruction. When deregulated, it is also critically involved in disease processes, notably in cancer, where protein kinases are now the major class of drug targets[94,95].

An essential step in understanding the complex molecular circuitry of cellular signal transmission is to develop methods for measuring the extent and nature of phosphorylation events that occur in a cell. For this purpose, modern quantitative mass spectrometry (MS) has proved to be an ideal platform because it is a highly precise yet generic method for the global identification and quantification of proteins and their modifications[96–99]. In eukaryotes, phosphorylation occurs almost exclusively on Ser (S), Thr (T), and Tyr (Y) residues, which represent approximately 17% of the total amino acids in an average human protein[100]. On this basis, there are nearly 700,000 different potential phosphorylation sites[101]. In 2003, MS-based studies identified ~300 phosphosites, and this was considered to be breakthrough progress at the time; but recent MS-based studies have reported the identification of tens of thousands of phosphorylation sites in tissues and cultured cells[102–104], and it has been speculated that the total extent of the phosphoproteome is more than a million[105]. This vast and increasing number of identified phosphosites raises

fundamental question about their properties and biological relevance and the scale of the complete phosphoproteome.

3.2.1. Phosphoproteomics in biomarker discovery

Phosphoproteomics has been extensively used as a preclinical research tool to characterize the phosphorylated components of the cancer proteome. For example, study done by Guo et al.[106] showed that growth factors and their downstream tyrosine kinases (e.g., Src), which are elevated during hormone-ablation therapy of prostate cancer, can induce tyrosine phosphorylation of androgen receptor (AR). Such modification may contribute to androgen-independent activation of AR or sensitize AR to low levels of hormone. The clinical importance of these findings is that inhibition of tyrosine kinases, such as Src that contribute to Y534 phosphorylation of AR, would be a more effective treatment regimen for hormone-refractory prostate cancer. Moreover, Wu et al, employed SILAC labeling based LC-MS/MS proteomic approaches to comprehensively profile the phosphoproteome of MCF7 breast cancer cells with tamoxifen resistance. Multiple crucial kinase-mediated signaling pathways were discovered to be hyperactivated in tamoxifen resistant cells, such as: MAPK, ERBB or insulin signaling pathways. Interestingly FAK2 tyrosine kinase was shown to play a pivotal role in development of tamoxifen resistance and potentially would be a very promising therapeutic target in context of resistance in breast cancer[107].

Furthermore, identification of phosphoproteins such as kinases as cancer biomarkers using MS-based phosphoproteomic technologies is an emerging field. Nowdays, most biomarkers used in clinical studies are based on proteomic applications as the majority of current drug targets are proteins, such as G protein-coupled receptors, enzymes and components of hormone signaling pathways[108] (93). They are critically important tools for detection, diagnosis, treatment, monitoring, and prognosis[109]. Importantly, cancer-specific phosphosignatures can be used as diagnostic markers and/or drug targets. For instance, study done by, Goss et al.[110] identified 188 Tyrphosphorylated sites, of which 77 were novel, from six cell lines representing three BCR-ABL fusion types and two distinct cellular backgrounds.

Comparison of this Bcr-Abl signature with the profile of cells expressing an alternative imatinib-sensitive fusion kinase, FIP1L1-PDGFRalpha, revealed that these kinases signal through different pathways. This study thus, revealed novel disease and potential drug-responsive biomarkers in the context of BCR-ABL kinase signaling, and demonstrated that phosphoproteomic approaches can be useful for discovering drug-responsive biomarkers.

In triple negative breast cancer, unfortunately, the field of biomarker development remains immature, without clinically relevant predictive biomarkers, highlighting the need for biomarkers to guide therapy in the future. Thus the important change must be made in this field with focus on simultaneous development of biomarkers with new therapeutic agents, beginning in the preclinical phase.

OBJECTIVES

Triple-negative breast cancer (TNBC) accounts for only 15% of the new breast cancer cases but kills up to 50% of the patients suffering from this diagnosis despite the best standards of care and early diagnosis. This relapse rate is unacceptably high compared with the other breast cancer subtypes (less than 20% in long-term series). Gene expression or histology approaches seem to distinguish several subtypes within TNBC, but no clear prognostic information is drawn out of such classifications. Moreover, these approaches do not provide potential therapeutic solutions or suggest novel targets to be explored. Unfortunately, recently published “cancer atlases” have not found sufficient oncogenic-addiction driving alterations at the gene-code, gene-expression or amplification/deletion level that could explain the inter-individual differences observed in TNBC patients population. Thus these data suggest that TNBC is not a single-alteration driven malignancy, but rather the result of a plethora of genetic, epigenetic, and microenvironmental alterations. Importantly, overall the tumour aberrations observed across the different layers of regulation of the tumour phenotype, eventually converge into discrete patterns of activation of signaling nodes. Thus we hypothesize that pinpointing these nodes would be a parsimonious solution for functional classification. Moreover such nodes, more specifically kinases, could also constitute potential druggable targets in TNBC. The objective of this thesis is to interrogate the phosphoproteome of triple-negative breast cancer cases from a functional point by addressing the following aims:

1. To perform high-throughput phosphoproteomic analysis using mass-spectrometry of triple-negative breast cancer cases in order to identify and quantify candidate phospho-proteins driving the aggressive nature of relapsed cases.
2. To validate phosphorylation signature driving the relapsed cases in a validation set of triple-negative breast cancer samples by immunohistochemistry.
3. To perform phosphoproteomic analysis of aggressive and indolent triple-negative cell lines in order to know whether their phosphor-profile mimics the profile found in relapsed vs. cured cases.
4. Target the phosphorylation signature found to drive relapsed cases in xenograft model of aggressive cell lines by pharmacological inhibitors.

OBJETIVOS

El cáncer de mama triple negativo (TNBC), representa sólo el 15% de los casos nuevos de cáncer de mama pero sigue teniendo un alto indice de mortalidad de hasta el 50%, a pesar de las mejoras en la atención y del diagnóstico precoz. Esta tasa de recaida es inaceptablemente alta en comparación con los otros subtipos de cáncer de mana (menos del 20% en series de pacientes a largo plazo). Estudios de expresión de genes o histología han identificado varios subtipos dentro del TNBC. A pesar de ésto, estas clasificaciones parecen carecer de importancia desde el punto de vista pronóstico de la enfermedad. Por otra parte, estos enfoques no proporcionan posibles soluciones terapéuticas ni sugieren nuevas dianas que puedan ser exploradas. Por desgracia, el articulo publicado recientemente titulado "Atlas de cáncer" no ha encontrado suficientes adicciónes oncogénicas que conduzcan alteraciones a nivel de genes, expresión de genes o amplificaciónes y deleciones, que podría explicar las diferencias interindividuales observadas en la población de pacientes de TNBC. Por lo tanto , estos datos sugieren que el TNBC no es una enfermedad conducida sólo por una alteración, sino más bien es el resultado de un panel de alteraciones geneticas, epigeniticas y microambientales. En general las aberraciones tumorales se observan a través de las diferentes capas de la regulación del fenotipo tumoral, eventualmente convergen en discretos patrones de activación de nodos de señalización. Por lo tanto presumimos que localizar estos nodos sería una solución útil para la clasificación funcional de este tipo de cáncer. Por otra parte tales nodos, más específicamente kinasas, también podrían constituir potenciales dianas terapéuticas para el tratamiento de TNBC.

El objetivo de esta tesis es interrogar el fosfoptoteoma de casos de cáncer de mama triple negativo desde un punto de vista funcional por abordar los siguientes objetivos:

1. Hacer un análisis de fosfoproteoma de alto rendimiento usando espectrometría de masas usando casos de cáncer de mama triple negativo con el fin de identificar y cuantificar fosfo-proteínas que conducen a la naturaleza agresiva de los casos que recaen.
2. Validar la señal de fosforilación que conduce a los casos de recaida en las muestras de pacientes de validación por inmunohistoquimica.

3. Llevar a cabo un análisis de fosfoproteómica en las líneas célulares triple negativas, tanto las que presentan un comportamiento agresivo como no agresivo con el fin de estudiar si presentan un patrón fosfoproteómico que mimetice el encontrado en los pacientes con o sin recaida.
4. Combatir el perfil de fosforilación encontrada en los casos que recaen en modelos de xenograft de líneas celulares con comportamiento agresivo utilizando los inhibidores farmacológicos.

MATERIALS AND METHODS

1. PATIENT SAMPLES

1.1. Samples for training and validation set

Samples from the training and validation set were gathered through the Spanish National Biobank Network. The study was conducted in accordance with the Declaration of Helsinki. Institutional review board approval was obtained at the Instituto de Salud Carlos III (study number: CEI PI 30-2010). For the training set 34 frozen tumor samples from patient diagnosed with triple-negative breast cancer were selected from three different Spanish hospitals (University Hospital Miguel Servet in Zaragoza, Gregorio Marañon Hospital in Madrid and University Hospital Complex Of Vigo) on the bases of biphasic relapse pattern (13 relapse within < 3 years and 21 non-relapsed after > 10 years) with similar distribution of conventional breast cancer prognostic variables like tumor size (T), grade (G) and nodal status (N). None of the patients received neoadjuvant therapy prior to biopsy or were BRCA-1 deficient. The validation set, 113 formalin-fixed and paraffin-embedded consecutive TNBC cases diagnosed in 2002 with triple-negative breast cancer were collected from two Spanish hospitals (MD Anderson and Hospital 12 de Octubre in Madrid).

2. CELL CULTURE

2.1. Triple negative breast cancer cell lines

A panel of Human Triple negative breast cancer cell lines (MDA-MB-231, MDA-MB-157, Hs578T, MDA-MB-468, HCC1937, HCC1143, HCC38, BT549, BT483 and MDA-MB-415) were maintained following the American Type Culture Collection recommendations (ATCC, Teddington, UK). Luminescent cell lines were generated by transfection with a plasmid encoding firefly luciferase (pGL.4.51 luciferase reporter vector, Pomega). Stable clones were isolated in vitro with Neomicin.

2.2. Cell proliferation assays

Breast cancer cell lines were seeded ($5x10^3$ cells/well) in 96-well plates for 24h and subsequently replaced with fresh medium with vehicle (control) or serial dilutions of the following drugs: 2.5-20 μM A12B4C3 (TOCRIS), 0.5-5 μM palbociclib (PD0332991) isethionate (Selleckchem), 5-40 μM GDC-0994 (MedChem, Princeton, NJ, USA), 2.5-20 μM imatinib (STI571) (Selleckhem), 1-10 μM LY2584702 tosylate (Selleckchem) and 2.5-25 μM of PKC Epsilon inhibitor peptide (Cayman Chemical, Ann Arbor, MI, USA) in 1x PBS. Inhibitory concentrations (IC50) values were determined as concentrations that inhibit 50% of cell viability. Viability assays were performed in triplicates, and replicates were normalized to cells treated with vehicle. For drug combination assays, $5x10^3$ cells/well were seeded in 96-well plates, left to attach for 24h and treated with IC50 of single compounds or combination of two compounds. After 72h cell viability was analyzed using CellTiter-Glo Luminescent Cell Viability Assay (Promega) following manufacturer instructions. For treatment combinations, the thresholds for antagonistic, additive, or synergistic effects were calculated as follows: DA = Da + Db (1 − Da), where Da is the relative inhibition (from 0 to 100%) value for the first drug and Db is the relative inhibition value for the second drug. Variations within ±15% of the additive threshold are considered to be the result of additive drug interactions; combinations exerting <85% of the predicted additive effect are considered indifferent (and below 70%, antagonistic); combinations exerting >115% of the predicted threshold for additive effect synergistic.

3. MICE STUDIES

3.1. Study approval

All animal experiments were approved by the CNIO Ethics Committee (PROEX/027/15) and performed in accordance with the guidelines stated in the International Guiding Principles for Biomedical Research Involving Animals developed by the Council for International Organizations of Medical Sciences. Four- to 6-week-old female athymic nude mice (Hsd: Athymic Nude-Foxn1nu) were purchased from Charles River Laboratories.

3.2. Animal models

For astringent model of metastasis $5x10^6$ luciferase expressing cancer cell lines were suspended in 200 µl of 1x PBS (Dulbecco) and injected intraperitoneally into female nude mice. Bioluminescence signal was followed once per week by IVIS Spectrum imaging system (PerkinElmer). For mammary fat pad injections $1x10^6$ cells were re-suspended in 50% Matrigel Basement Membrane Matrix (Corning). Tumor formation and growth were monitored weekly by using calipers. Tumor volumes were calculated using the formula *V* = (*D* °— *d*2)/2 mm3, where *D* is the largest diameter and *d* is the shortest diameter. Mice were euthanized when reaching human end point. Tumors were excised and fixed (10% formalin solution) for histological examination (FFPE) or snap-frozen for subsequent analysis.

3.3. Treatments

In case of intramammary injections when the tumors reached $500mm^3$ size, mice were randomized (http://www.randomization.com/) to receive either single compound or combinations. For intraperitoneal injections mice were randomized to receive the same treatment regimens two days after the injection. For the animal treatments palbociclib, dissolved in 50mM sodium lactate buffer, pH 4.0, was administered by oral gavage at 100mg/kg/day. GDC-0994 was dissolved in 2% DMSO+ 30% PEG-300+ 5% Tween80+dH2O and administered by oral gavage at the dose of 50mg/kg/. Imatinib was dissolved in sterile saline and administered by intraperitoneal injection at the dose of 70mg/kg/day.

Tumor growth inhibition (TGI) was calculated using the following formula: TGI = [1 − (*T*F/*T*0)A/(*T*F/*T*0)V] × 100, where *T*F is the time point analyzed, *T*0 is the initial time, A () is the tumor measurement corresponding to drug treatment , and V() is the tumor measurement from the vehicle treatment.

For treatment combinations, the thresholds for antagonistic, additive, or synergistic effects were calculated as follows: TA = Fa + Fb (1 − Fa), where Fa is the relative TGI (from 0 to 1) value at a given time point for the first drug and Fb is the relative TGI value at the same time point for the second drug. Variations within ±15% of

the additive threshold are considered to be the result of additive drug interactions; combinations exerting <85% of the predicted additive effect are considered indifferent (and below 70%, antagonistic); combinations exerting >115% of the predicted threshold for additive effect synergistic. The effects of the combinations are calculated at the last time point when the animals belonging to the treatment group that was first terminated because of tumor growth among those compared were alive.

3.4. Generation of PNKP knockout cells

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 All-in One Lentivector system was purchased from Applied Biological Materials (ABM). Recombinant lentiviruses were generated in HEK 293T cells by cotransfection of sgRNA-encoding plasmids targeting human PNKP (Cat#K1676005) with pVSV-G and psPAX2 packaging plasmids (Addgene). Stable integrants were selected with puromycin.

4. PHOSPHO-PROTEOMIC ANALYSIS

All tumor samples in study had more than 75% tumor content. We interrogated the phosphoproteome by enriching the phospho-fraction out of 250 micrograms of purified protein per sample by Ti4-IMAC chromatography coupled with mass-spectrometry based on previous results, where we showed that this technique outperformed TiO2 and IMAC, other commonly used phospho-enrichment strategies[111]. In addition, given the large number of samples and the limited amount of material of the clinical specimens, we aimed to probe the phosphoproteome using a label-free single-shot strategy as described previously by de Graaf et al[112]. Such approach allows highly reproducible measurements, both qualitatively and quantitatively, over a large dynamic range[112]..

One (12 samples), two (19 samples) or three (3 samples) phospho-enrichments were performed in the tumor samples depending on the available starting material; two and three phospho-enrichments were performed in 7 and 3 cell lines, respectively. Each

sample was run twice, for a total number of 118 (tumors) and 46 (cell lines) runs; The average phosphoenrichment ratio was 93.3%.

4.1. Protein extraction from frozen patient tumor biopsies and triple-negative cell lines

Frozen tumor samples were homogenized using a Precellys 24 device (Bertin Technologies, Paris, France) in ice-cold lysis buffer containing 7 M urea (ThermoFisher Scientific), 2 M thiourea, 2 % N-octyl glucoside (Santa Cruz Biotechnology, Inc.), 15 mM tris (2-carboxyethyl) phosphine (TCEP), 50 mM Hepes (pH 7.5), protease and phosphatase inhibitor cocktail (Halt EDTA-free; Thermo Scientific; Waltham, MA, USA), and 0.1% Benzonase Nuclease (Novagen, Darmstadt, Germany).

Tumor cell lines were washed 2 times with cold PBS 1x (Dulbecco´s, Sigma-Aldich) and harvested in ice-cold lysis buffer containing 7 M urea (ThermoFisher Scientific), 2 M thiourea, 2 % N-octyl glucoside (Santa Cruz Biotechnology, Inc.), 15 mM tris (2-carboxyethyl) phosphine (TCEP), 50 mM Hepes (pH 7.5), protease and phosphatase inhibitor cocktail (Halt EDTA-free; Thermo Scientific; Waltham, MA, USA), and 0.1% Benzonase Nuclease (Novagen, Darmstadt, Germany).

Protein lysates from tumors and cells were sonicated for 15 min on ice and clarified by centrifugation at 14,100×g at 4°C for 15 min. Protein concentration was estimated by a colorimetric assay (660 nm protein assay; Pierce; Rockford, IL, USA) using bovine serum albumin as reference.

4.2. Protein digestion, sample clean-up and Ti4+-IMAC phosphopeptide enrichment

Protein samples were digested using the filter-aided sample preparation method[113]. Briefly, 250 µg of each tumor and cell line protein extract dissolved in lysis buffer was reduced with 10 mM DTT for 20 min at 56°C and alkylated using 50 mM IAA soloution for 20 min at 25°C in the dark. The excess of reduction and alkylation reagents were washed with 8 M urea in 100 mM Tris (pH 8). The proteins were digested

at 37°C in wet chamber for 4 h using endoproteinase Lys-C at a 1:50 enzyme-to-protein ratio. After Lys-C digestion, urea was adjusted to 2M with 50 mM triethyl ammonium bicarbonate. Trypsin (Promega) was added at a 1:50 (enzyme-to-protein ratio) and samples were subjected to a second digestion at 37°C overnight in wet chamber. The digestion was quenched by acidification with formic acid (0.1% final concentration). Prior to phosphopeptide enrichment, digested samples were cleaned up with reversed phase-based solid-phase extraction (SPE). Phosphopeptide enrichment was performed using $Ti4^{+}$-IMAC GELoader spin tips by centrifugation[114]. Briefly, Geloader tip microcolumns packed with Ti4+-IMAC beads were created using a C8 plug at the constricted end. An aliquot of the Ti4+-IMAC beads suspension was packed in the tip at the ratio 2:1 of beads:peptides. Max 250µg of each digested tumor or cell line sample was loaded onto the column and centrifuged at 50g for 30min to allow the binding of phosphopeptides to the Ti4+-IMAC beads (~3 µl / min). Bound phosphopeptides were washed with 80% acetonitrile (ACN) in 6% trifluoroacetic acid (TFA) by centrifugation at 170g (~3 µl / min). Afterwards two additional washing step were performed, first one with 50% ACN/ 0.5%TFA in 200mM NaCl and second one with 50% ACN/0.1% TFA. Both washing steps were performed by centrifugation at 170g (~3 µl / min). Phosphopeptides were eluted from the column with 10% ammonia into 25% formic acid (FA) and centrifugated at 100g (~1 µl / min). Finally, a second elution was carried out with 80% ACN/2%FA and centrifuged at 100g (~1 µl / min). Phosphopeptides were acidified by adding 3µl of 100% FA. The eluate was directly injected and analyzed by LC-MS/MS.

4.3. Reverse Phase Chromatography and Mass-spectrometry

Peptides were subjected to reverse phase nano-LC-MS/MS analysis using a Proxeon EASY-nLC 1000 (Thermo Scientific, Odense, Denmark) with an analytical column heater (40 °C) and an LTQ-Orbitrap Elite (Thermo Fisher Scientific, Bremen, Germany). Peptides were first trapped (Reprosil C18, Dr Maisch, GmbH, Ammerbuch, Germany, 3 µm, 2 cm × 100 µm) at a maximum pressure of 800 bar with 100% solvent A (0.1% formic acid in water) before being separated on a 40cm x 50µm analytical column (Poroshell 120 EC-C18, 2.7µm, Agilent, Santa Clara, CA). Peptides were

chromatographically separated by a 90-min gradient from 7% to 30% solvent B (0.1% formic acid in ACN) at a flow rate of 100 nL/min. The total measurement time for each sample was 120 min. The eluent was sprayed via a distal coated fused silica emitter (360-µm outer diameter, 20-µm inner diameter, 10-µm tip inner diameter; constructed in-house) butt-connected to the analytical column. The electrospray voltage was set at 1.7 kV. The mass spectrometer was operated in a data-dependent mode to automatically switch between MS and MS/MS. Briefly, survey full-scan MS spectra were acquired in the Orbitrap analyzer, scanning from *m*/*z* 350 to *m*/*z* 1500 at a resolution of 60,000 at *m*/*z* 400 using an automatic gain control setting of 1e6 ions. Charge state screening was enabled, and precursors with either unknown or 1+ charge states were excluded. After the survey scan, the 20 most intense precursors were selected for subsequent decision-tree-based ion trap CID or ETD fragmentation[115]. The normalized collision energy for CID was set at 35%, and supplemental activation for ETD and dynamic exclusion were enabled (exclusion size list: 500; exclusion duration: 40 s).

4.4. Data analysis and Kinase prediction

The raw MS data were processed with MaxQuant software suite version 1.3.0.5[116] The fragmentation spectra were searched against the *Homo sapiens* Uniprot database (downloaded on 23-12-2013), using Andromeda as the search engine. The precursor mass tolerances were set to 20 ppm for the first search and 4.5 for the main search. Also, 0.05 and 0.5 Da were used for FT and IT detectors. Carbamidomethylation of cysteine was considered as fixed modifications, whereas oxidation of methionine (M); phosphorylation on serine (S), threonine (T) and tyrosine (Y); and protein N-terminal acetylation were chosen as a variable modification, and up to two tryptic missed cleavages were allowed. The match between run function was enabled. A target-decoy database searching strategy was used to evaluate the false-discovery rates (FDRs) at the peptide and protein level.

The identification of kinase specific substrates was evaluated using linear sequence motifs analysis implemented in MaxQuant[116]. For the identification of phosphorylated motifs Maxquant used the PhosphoMotif Finder search tool at Human Protein Reference Database was used (http://www.hprd.org/PhosphoMotif_finder)[117].

4.5. Consensus clustering

We applied the Consensus Clustering module using ConsensusClusterPlus package available in Bioconductor (http://www.bioconductor.org/) for class discovery and clustering validation. This method facilitates the discovery of biologically meaningful clusters assessing the stability for the discovered clusters by means of resampling techniques. Briefly, a clustering method chosen by the user is applied to each of the resampled data sets and, the consensus among the multiple runs, is assessed and summarized in a consensus matrix. This matrix is used as a visualization tool to estimate the cluster composition and number. For this study the Consensus Clustering analyses were run with hierarchical clustering algorithm, and Pearson correlation using 500 resampling iterations.

4.6. GO pathways

Gene-ontology pathway analysis was performed with Enrichr: a comprehensive gene set enrichment analysis (http://amp.pharm.mssm.edu/Enrichr/#)[118,119].

4.7. Differential analysis

Differentially expressed phosphopeptides were obtained by applying linear models with R limma package[120]. (Bioconductor project, http://www.bioconductor.org). To account for multiple hypotheses testing, the estimated significance level (*p* value) was adjusted using Benjamini & Hochberg False Discovery Rate (FDR) correction. Those phospho-sites with FDR <0.15 were selected as differentially phosphorylated between classes under comparison. Representative Heat-Map was depictured using GENE-E software. (www.broadinstitute.org/cancer/software/GENE-E/)

4.8. Enrichment analysis

Gene Set Enrichment Analysis (GSEA) was used to define set of kinase substrate motifs that shows statistically significant, concordant differences between relapsed and cured cases and aggressive and indolent triple-negative cell lines. To this end, GSEA was applied using annotations for motifs extracted from Perseus software.

Leading proteins were ranked based on their intensities. After Kolmogorov-Smirnoff testing, those gene sets showing FDR <0.05 were selected as significant[121].

5. IMMUNOHISTOCHEMISTRY

5.1. Preparation of tissue microarrays (TMAs)

Tissue microarrays were mounted with 2 1-mm cores per sample (MTA-I, Beecher Instruments), including only the 113 samples from the validation set. An expert pathologist examined a template H&E slide from each sample to select the areas for core selection.

5.2. Staining of TMAs with specific antibodies

Immunohistochemical staining was performed on 5-µm TMA sections. Deparafinization and antigen retrieval (cell conditioning) were performed on DISCOVERY XT automated slide staining system using validated reagents (Ventana Medical Systems, Inc.,).

The following antibodies were used for IHC: phospho-AKT (Ser473) (D9E), phospho-p70S6K (Thr389) (1A5), phospho-p44/42 (ERK1/2) (Thr202/Tyr204), phospho-S6 Ribosomal Protein (Ser240/Ser244) (D68F8) and phosho-PNKP (Ser114/Thr118) (all from Cell Signaling Technology); CDK6 (98D/H8, Monoclonal Antibodies Core Unit, CNIO); PKC Epsilon [(EPR1482(2)]; PIM1 (12H8, Santa Cruz Biotechnology, Inc.), c-KIT/CD117 (Dako), CLK-1 (Sigma-Aldrich) and DAPK3/ZIPK (LifeSpan BioScience).

Corresponding TMA were acquired and digitalized using the Ariol SL-50 system coupled with fully automated microscope system Leica DM6000 B. Scores were generated using the TMASight assay, providing areas of high staining (Area_Color 1), medium staining (Area_Color 2) and low staining (Area_Color 3). We calculated the percentage of each staining per biopsy normalized to Area_Color 4, which represents the whole tissue, providing a computerized H- score calculated by formula:((% of Area_Color1*3) + (% of Area_Color2*2) + (% of Area_Color 1*1))/100

5.3. Immunoblots

Cells were washed 2x with PBS and harvested in cold RIPA Buffer (Sigma) containing 1% protease and phosphatase inhibitor cocktail (Halt EDTA-free; Thermo Scientific). Cell lysates were incubated at 4°C for 15 min, sonicated for 15 min and clarified by centrifugation at 14 000xg at 4°C for 30 min. Protein concentration was estimated by a colorimetric assay (660 nm protein assay; Pierce) following the manufacter's instruction. 20 µg of proteins per sample were loaded on 10% SDS-PAGE gel and transferred to nitrocellulose membranes for further processing. 5% BSA was used to block the membrane for 60 min at room temperature, followed by overnight incubation at 4ºc with the primary antibodies.

The following primary antibodies were used: **phospho-p70S6K** (Thr389), **p70S6K phospho-p44/42** (ERK1/2) (Thr202/Tyr204), **p44/42** (ERK1/2) and **phosho-PNKP** (Ser114/Thr118) (all from Cell Signaling Technology); C**DK6** ready to use (clone 98D/H8, Monoclonal Antibodies Core Unit, CNIO), **PKC Epsilon** (EPR1482(2), LifeSpan BioScience), **c-KIT/CD117** (clone YR145, Milipore), **PNKP** (Novibio), Vinculin and β-Actin (clone AC-15; both from Sigma- Aldrich). Membranes were incubated with appropriate peroxidase-conjugate secondary antibodies (Sigma, St Louis, MO, USA). Bands were visualized by the enhanced chemiluminescence (ECL) method (Lumi-LightPlus detection kit; Roche).

6. STATISTICAL ANALYSIS

The prognostic value of each candidate kinase was calculated in two ways: with the Kaplan Meier method and the Log-Rank test (univariate), and with the Cox´s proportionate hazards model (multivariate model, adjusting the hazard ratio attributable to each kinase by the T and N status, age and grade), implementing a multiple comparisons correction to the cut-off point of $P<0.05$. The kinase scores were categorized: patients with a kinase staining above the 75^{th} percentile were encoded as "1", whereas the remainder were encoded as "0" (both for the Kaplan-Meier tests and the Cox models). The kinases with positive association with relapse were grouped into a

single variant (“K-high”); patients with a staining in one or more (positively associated) kinases above the 75th percentile were encoded as K-high=1; in order to be encoded as K-high=0, patients had to have a staining below the 75th percentile in all six kinases. The Kaplan-Meier method and the Cox´s proportionate hazards model were equally applied to investigate the K-high association with relapse (adjusted by age, grade, T and N status). Kinase co-linearity was investigated with the Pearson´s coefficient in a pairwise manner.

Tumor burden comparisons and survival benefit between treatment groups for the in vivo experiments were performed with a T-test (average burden obtained from at least 10 tumors per time point and condition) and the Kaplan-Meier method (used to compare the time-to-sacrifice between treatment groups). All tests were performed with the SPSS Statistics V.19.0 software.

RESULTS

1. CHARACTERIZATION OF PHOSPHO-PROTEOMIC LANDSCAPE IN TRIPLE-NEGATIVE BREAST CANCER

Profiling breast cancer with expression arrays has become common, and it is believed that such studies will allow better understanding of the molecular differences between clinical cases, thus improving individualization of care[122]. However, only a partial concordance between genomic copy number, RNA levels and protein levels in both patient samples and cell lines, has been reported[123–125]. This can be explained, at least, in part, because protein levels and, in particular, phosphoprotein levels represent an integration of the complex genomic and transcriptomic aberrations accumulated in each tumour combined with translational and post-transcriptional regulation that cannot be fully captured by genomic and transcriptomic analysis. Protein phosphorylation is of critical importance in signalling systems and employed by cells to transiently alter different protein properties such as: protein-protein interactions, activity or degradation. When deregulated, it is also critically involved in disease processes, notably in cancer, where protein kinases are now the major class of drug targets[94,95].

In order to understand the complex network of signal transmission in triple-negative breast cancer, it was important to develop methods that could measure the nature and extent of phosphorylation events that occur in particular cell. For this purpose, the ideal platform has provided to be modern mass spectrometry (MS). This technique allows highly precise, yet generic method for global identification and quantification of proteins and their modifications[96–99]. In line with this, recent mass-spectrometry based studies were able to identify more than tens of thousands of phosphorylation sites in tissues and cultured cells[102–105]. This vast and increasing number of identified phosphosites raises fundamental question about their properties and biological relevance and the scale of the complete phosphoproteome.

The present thesis is focused on the study of phospho-proteomic landscape of triple-negative breast cancer patient samples and cell lines with the aim to discover novel prognostic and predictive markers that could better explain different clinical

outcome of this aggressive subtype and moreover set the stage for the discovery of novel alternative targeted therapies.

Due to the dynamic nature of the phosphoproteome, the analysis of clinical samples requires prudence. Several pre-analytical variables need to be considered when investigating phosphorylation changes within human specimens, especially when molecular analysis is incorporated into research and clinical practice to investigate prognostic/predictive signatures in biological material. Firstly, the signaling network is a highly dynamic system and it is deeply influenced by environmental factors such as stress, hypoxia and nutrient supply, thus appropriate preservation of the biological material is absolutely necessary for obtaining accurate and reliable measurements of intracellular signaling[126]. Secondly, the overall amount of biological material normally available for molecular profiling is relatively small (for example, fine needle aspirates/core biopsies) and cell subpopulations of interest usually represent small percentages of the whole tissue prior to upfront cellular enrichment. For this reason, the proteomic platform intended to be used for clinical and translational studies, has to provide high-throughput, accurate and reproducible results, even from relatively little starting material, such as few thousand malignant cells.

1.1. Selection of technical approach for the phospho-proteomic profiling of TNBC cases

Taking into account the importance and limited amount of biological samples we had available, together with desired coverage of phosopho-proteome, in the beginning of our study, it was essential to select the adequate technical approach to be used to obtain the phosphoproteomic signature of triple-negative breast cancer patient samples and cell lines. Moreover, it was important to determine the minimum amount of whole protein extract that we would need for good coverage.

In addition, large scale phosphoproteomics strategies typically include enrichments of phosphopeptides due to their low abundance, using immobilized metal affinity

chromatography (IMAC) or titanium dioxide (TiO_2) resin, followed by reverse phase LC coupled to electrospray ionization (ESI) tandem MS[127]. Thus, since previous studies[128] showed that Ti^{4+}-IMAC performed better in terms of the number of identified phosphoproteins, we decided to use this approach as phosphopeptide enrichment strategy. In line with this, to test the performance of Ti^{4+}-IMAC in samples from breast cancer patient biopsies and cell lines and to determine the minimal amount of protein extract we would need, we did phosphopeptide enrichment and subsequent mass-spectrometry analysis testing different amount of protein extract as a starting material. Additionally, we tested the performance of enrichment in protein extract from HeLa cells and MMTV-PyMT mouse tumor sample. Using this approach we were able to identify different number of phospho-peptides as summarized in Figure 1.

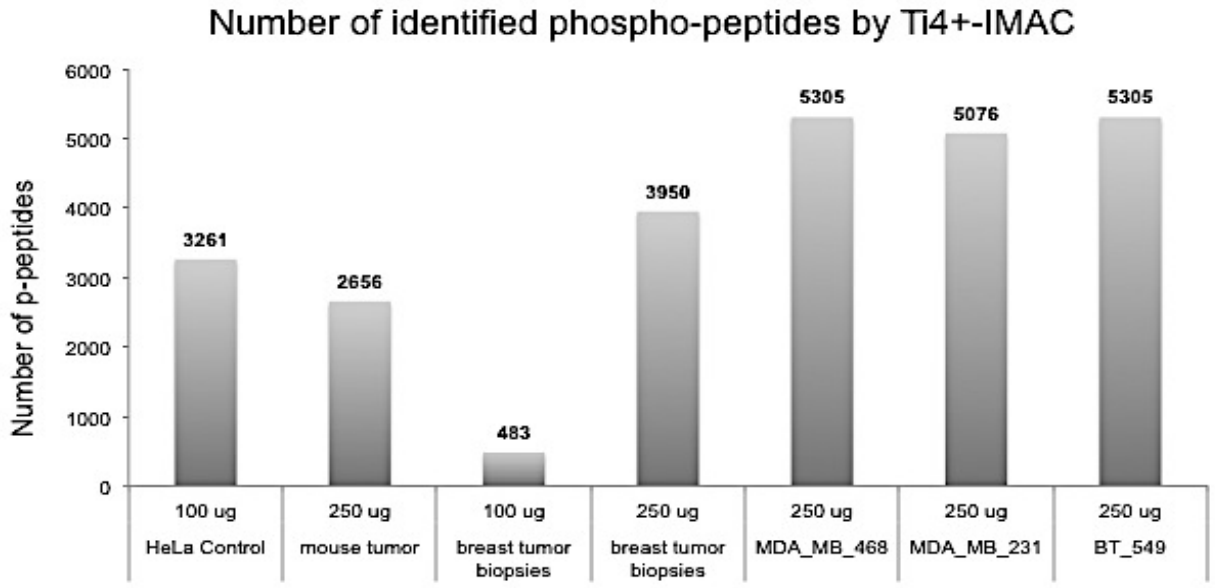


Figure 1. Number of phosphopeptides identified in samples from different origin and starting from different amount of protein lysate. 100µg of protein lysate from HeLa, 250µg of protein lysate from MMTV-PyMT mouse tumour, 100µg and 250µg of protein lysate from human breast cancer tumour and 250µg of protein lysate from three triple-negative breast cancer cell lines (MDA-MB-468, MDA-MB-231 and BT-549) were subjected to phosphopeptide enrichment using Ti^{4+}-IMAC, followed by LS-MS analysis. The graft represents the number of unique phosphopeptide identifications.

1.2. *Identification of phosphoproteomic signature in TNBC patient samples and cell lines*

In order to identify phosphoproteomic signature of triple negative breast cancer we firstly performed phosphopeptide enrichment and subsequent LC-MS analysis and label-free quantification of phosphoprotein dynamics using frozen tumour biopsies from TNBC patients. Moreover, in our phosphoproteomic profiling we included ten triple negative breast cancer cell lines in order to perform therapeutic experiments, later on, based on their phosphoprofile.

1.2.1. Phosphoproteomic profile of breast cancer cases

For phosphoproteomic profiling of TNBC cases we collected a training set of 34 frozen tumor biopsies showing biphasic pattern of relapse of the disease. This training set included two groups of patients. One group with TNBC patients who relapsed in less than 3 years (Group A) and the other group with patients that were relapse-free for more than 12 years (Group B). Patients in relapsed and non-relapsed group were paired by classical prognostic factors (Table 1, Figure 2).

	Group A (N=13)	Group B (N=21)
Age (mean, range)	54.1 (33-71)	58. 2 (41-77)
TNM Stage		
I	2 (15%)	4 (19%)
II	9 (69%)	13 (62%)
III	2 (15%)	4 (19%)
Type of chemotherapy		
Anthracyclines plus cyclophosphamide	3 (23%)	4 (19%)
Antracyclines plus cyclophosphamide plus taxanes	10 (77%)	17 (81%)
Ki67 average*	39%	42%
Grade		
G1	0 (0%)	0 (0%)
G2	5 (38%)	4 (19%)
G3	8 (61%)	17 (81%)
Premenopausal	9 (70%)	17 (81%)

*Ki67 status is not a prognostic factor in TNBC[129]

Table 1. Clinical and demographic characteristic of the training set

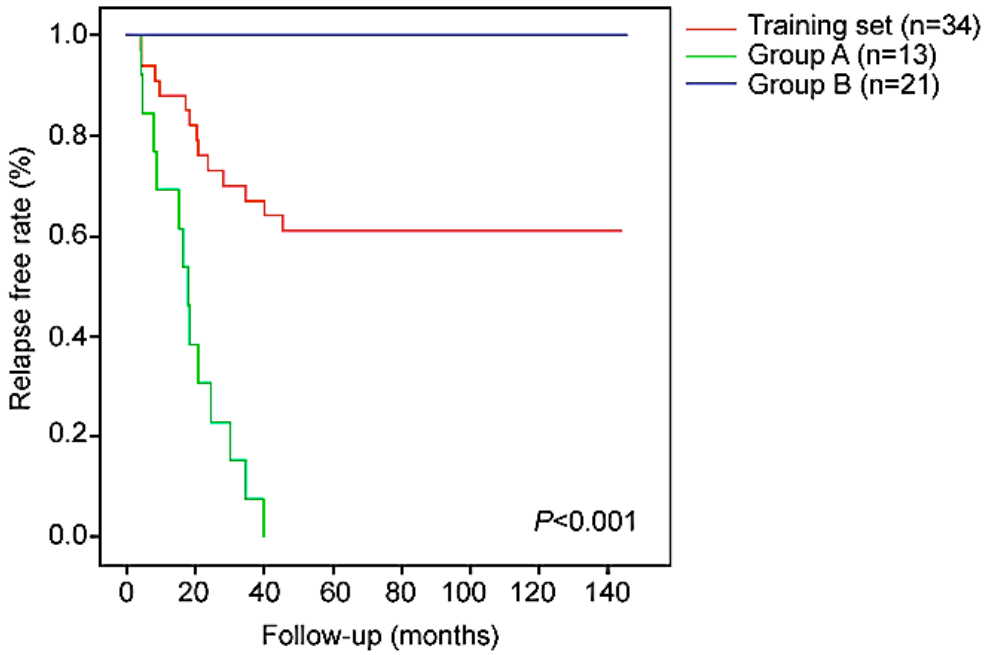


Figure 2. Two sub-groups of patients were involved in training set. Group A with 13 patients that relapsed within < 3 years of logo-regional treatment. Group B with 21 patients non-relapsed after > 10 years of logo-regional treatment. Kaplan-Meier curves of the training set (red) and its two subgroups.

Median time to relapse of group A: 17.36 months; median time to relapse of groups B: not reached (Log Rank $P < 0.001$)

Our pilot study (Figure 1), in which we tested the performance of Ti^{4+}-IMAC in breast cancer samples, clearly demonstrated that we were able to achieve a good number of identified phosphopeptides. However, it was crucial to have enough starting tissue material, such as no less than 250 µg of protein extract. Since, we got enough protein extract from all 34 frozen tumor biopsies we proceeded with Ti^{4+}-IMAC phosphopeptide enrichment of all protein digests and subsequently performed LC-MS analysis and label-free quantification. As shown in Figure 3, different number of phosphopeptides were identified in breast cancer cases.

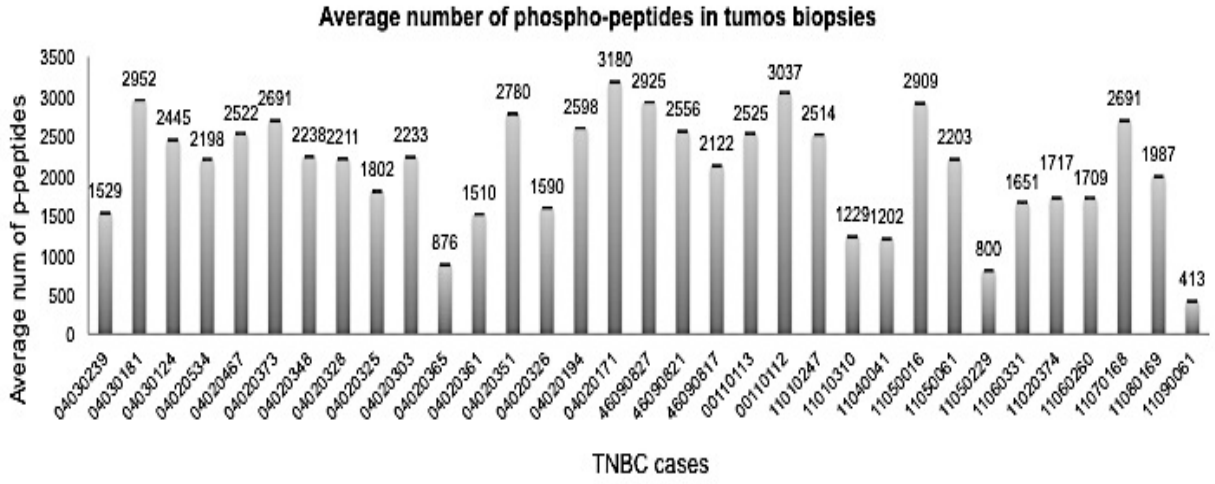


Figure 3. Average number of identified phosphopeptides in TNBC cases. Ti^{4+}-IMAC was performed on up to minimum of 250µg of protein extract. From each breast cancer frozen tumour we performed minimum one enrichment. For those samples that yield more than 250µg of protein extract we performed replicate of enrichment. Each enriched sample was injected at least two times in MS. Data are presented as average number of phosphopeptides in all the injections from each sample individually

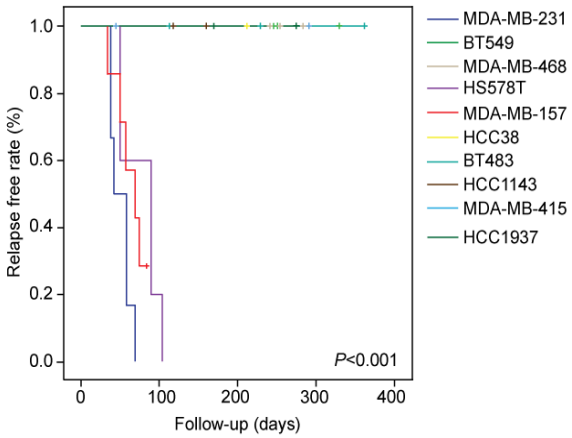
1.2.2. *Phosphoproteomic profile of triple-negative cell lines*

Since the aim of our study is functional analysis and validation of potential therapeutic targets in animal models, together with triple negative cases, we also phospho-profiled ten triple negative breast cancer cell lines. Prior to phosphoproteomic analysis we first characterized these cell lines for their in vivo aggressive potential. For

this purpose ten TNBC cell lines engineered for luciferase expression were injected intraperitoneally into Nude mice and their survival was followed for more than 400 days.

Our in vivo study showed that TNBC cell lines behaved as well in biphasic manner. Three cell lines, MDA-MB-231, MDA-MB-157 and Hs578T, showed more aggressive behaviour since were able to kill 100% of animals in less than 100 days. However the remaining seven TNBC cell lines showed more indolent behaviour since animals were tumour free for more than 400 days (Figure 4).

A



B

TNBC CELL LINE	MEDIAN TIME TO DEATH (days)
MDA-MB-231	50
MDA-MB-157	69
Hs578T	90
MDA-MB-468	not reached at 254
HCC1937	not reached at 275
HCC1143	not reached at 160
HCC38	not reached at 362
BT549	not reached at 251
BT483	not reached at 229
MDA-MB-415	not reached at 291

C

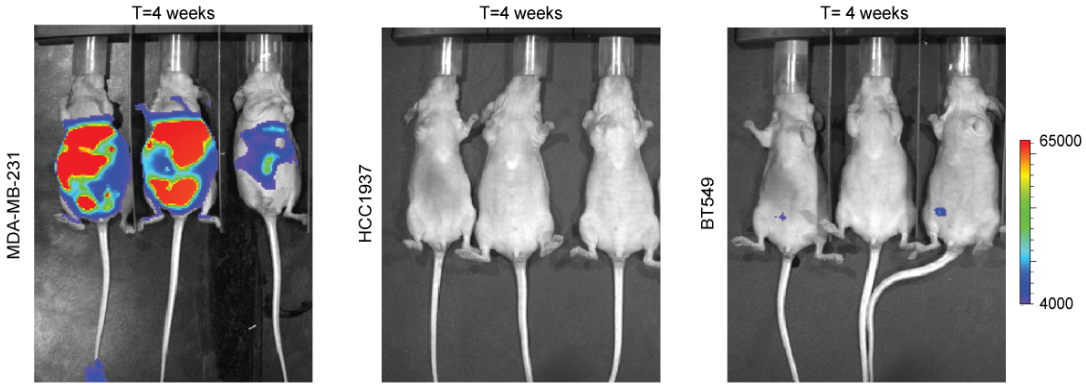


Figure 4. Summary of in vivo studies. (A) Kaplan-Meier curve of the survival of ice (N=60) injected with different TNBC cell lines intraperitoneally. The intraperitoneal injection is an astringent model of metastasis compared to tail-vein injection: the latter, compared to the former, skips the process of tumor cells leaving the primary site and entering the bloodstream. (B) Median time to sacrifice of animals groups injected with each cell lines (Log rank *P*< 0.001). (C) Representative images of tumor burden and dissemination in animals 4 weeks after injection of one of aggressive TNBC cell lines (MDA-MD-231) or two of the indolent cell lines (HCC1937 and BT549).

We were able to identify different aggressive phenotype in TNBC cell lines that would potentially mimic the aggressive behaviour seen in TNBC cases from our discovery set.

Next we did protein extraction from cell lines and subsequently performed phosphopeptide enrichment, LS-MS analysis and label-free quantification of phosphoproteomic dynamics. As shown in Figure 5 performing Ti^{4+}-IMAC enrichment the number of phosphopeptides that were identified in different cell lines was very similar (approximately 4000 phosphopeptides)

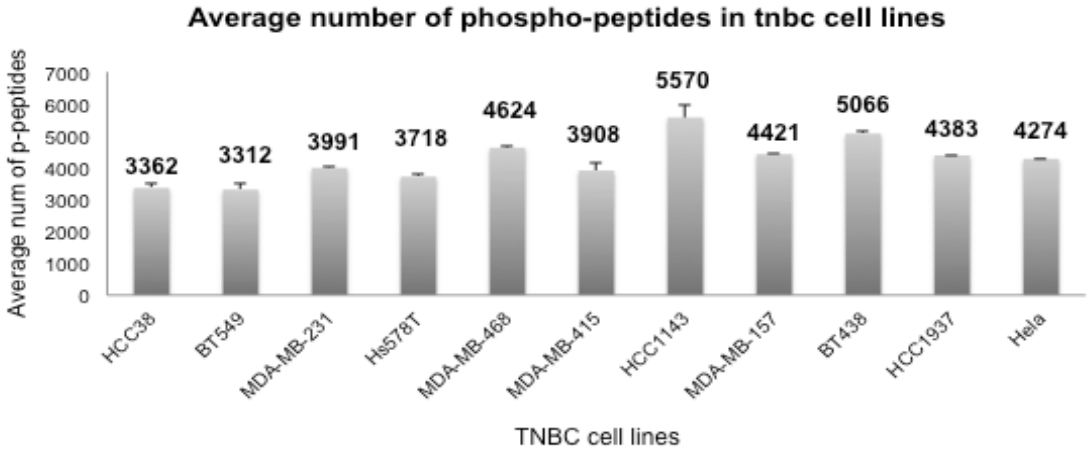


Figure 5. Bar chart representing number of identified phosphopeptides in TNBC cell lines and HeLa control sample. Ti^{4+}-IMAC was performed on 250µg of protein extract. From each cell line we performed two or three enrichment experiments (replicates). For each enriched sample two injections in MS were performed. Data are presented as average number of phosphopeptides in all the injections belonging to one sample.

1.3. Identification of differently phosphorylated peptides

After performing phosphoproteomic analysis in both cell lines and tumour biopsis, in order to weight how inherently different were the cell lines-derived phosphoprofiles from those originated in tumor biopsies we performed consensus clustering analysis (Figure 6). This bioinformatics analysis involves k-means clustering resampling (x1000) of randomly selected profiles (from either the first or second mass spectrometry runs, from cell lines or tumors). The upper panel (Figure 6A) shows the visual representation of the relative number of times in which 2 profiles, cluster together during those 1000

iterations. The darker the colour is the more frequently two profiles cluster next to each other. In order to estimate the “real” number of clusters present in the phospho-matix, we plotted the area under the curve (Delta) of the consensus distribution function (CDF). This function shows that the optimum number of clusters is two, with a proficient separation between cell lines and samples (Figure 6B,C).

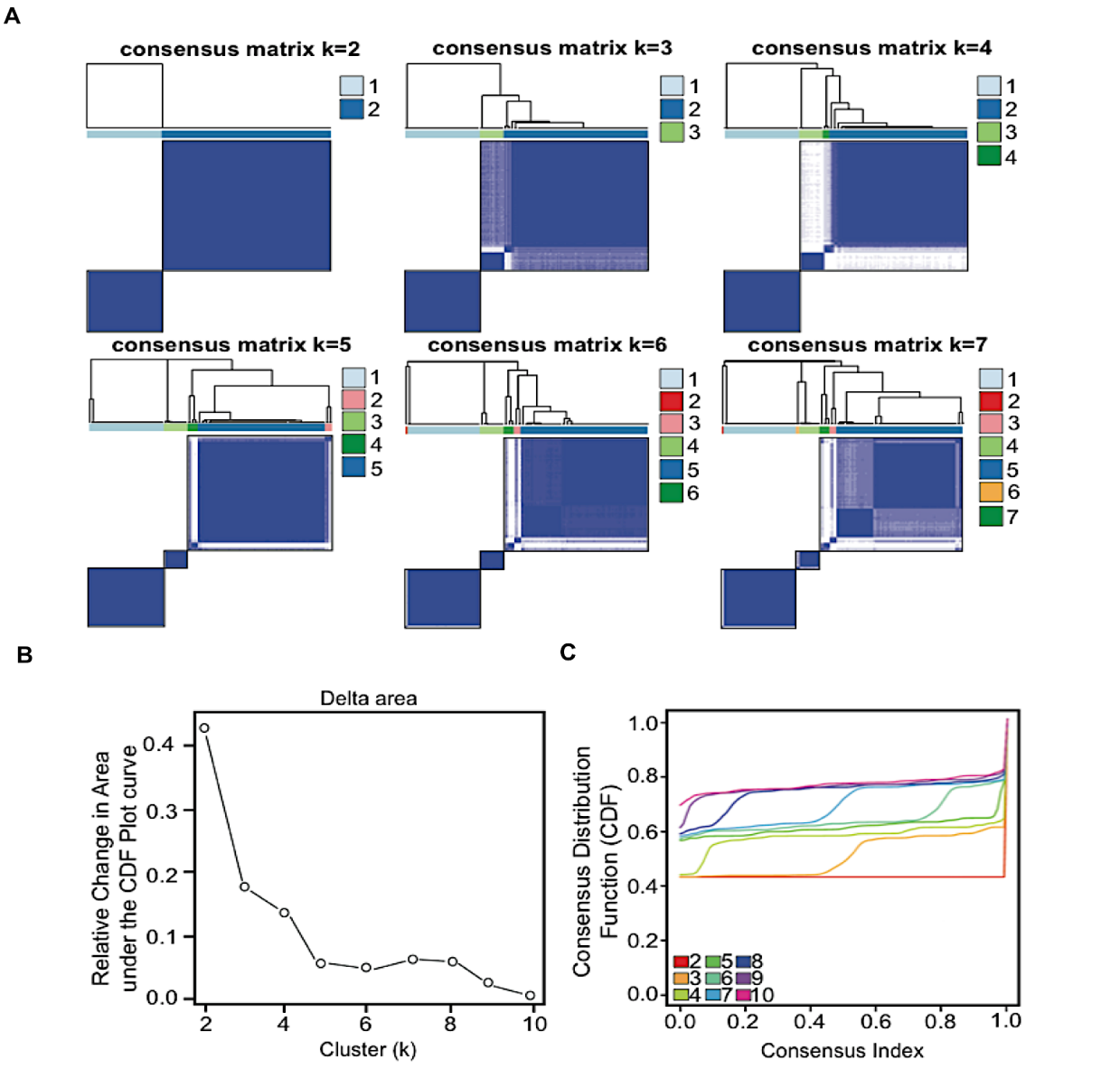


Figure 6. A consensus clustering analysis of phosphoprofiles of TNBC tumors and cell lines. (A) A consensus clustering analysis was applied to weight the difference between the phosphoprofiles of the cell lineas and the phosphoprofiles of the tumour samples. In the figure A is the visual representation of almost perfect separation of two phosphoprofiles. (B)(C) As indicated in the plots the real number of clusters is two, which is estimated when delta area under the CDF curve is plotted for each k-means.

The observed difference between tumor biopsies and cell lines is highlighted in the Venn diagram shown in Figure 7, where approximatelly 30% of phospho-peptides were commonly identified in tumor biopsies and cell lines. From the proteins to which these phospho-peptides mapped, we performed gene-ontology (GO) analysis in order to find pathways enriched in each group. We found important overlap in GO pathways, even from non-redundant phospho-peptides, with only few pathways unique to each group (exp Tight junction, RNA transport).

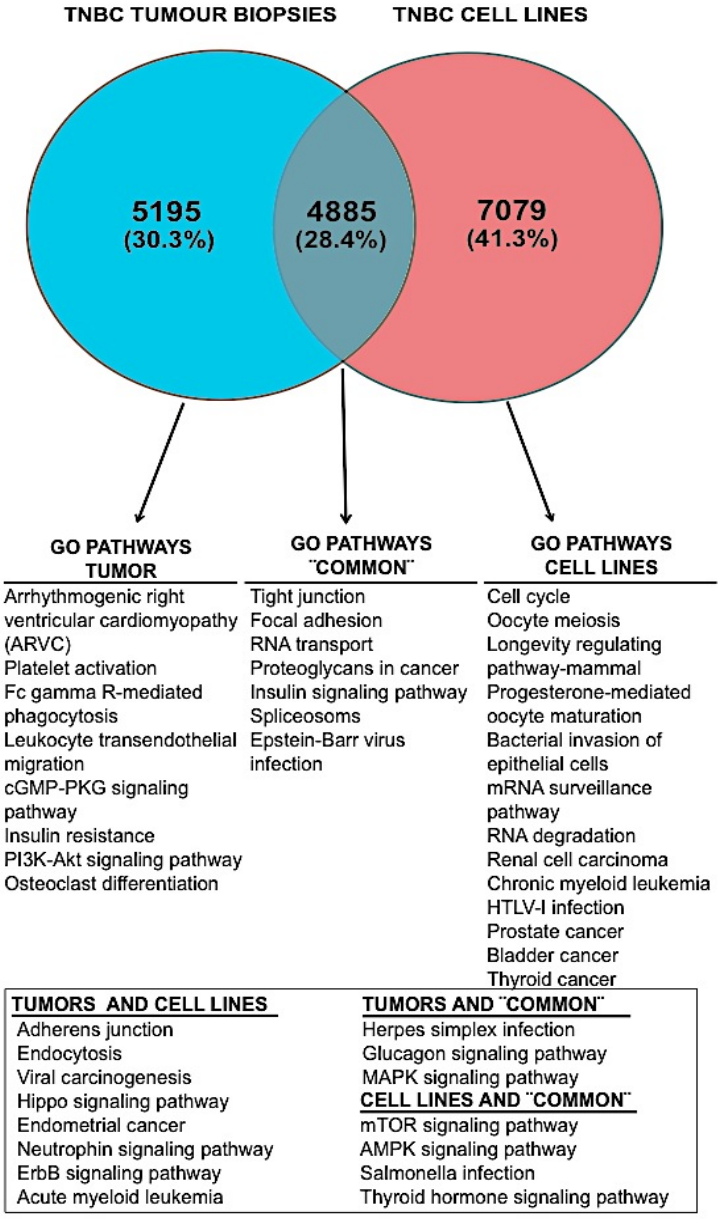


Figure 7. Venny diagram of phosphopeptides and gene ontology (GO) pathways significantly enriched in tumour biopsies, cell lines or both (P<0.05).

Even though we found interesting and important pathways in each group of samples, one important issue that should be highlighted is that we currently lack good bioinformatic tools that could integrate the status of phospho-proteins into higher level of information beyond GO pathways. Thus, to overcome this, we looked for differently phosphorylated peptides between relapsed and non-relapse patients and aggressive and indolent cell lines by performing differential analysis. First the log2 normalized phosphopeptide data from patient samples and cell lines were filtered to include phosphopeptides with signal in at least one sample/replicate. This analysis gave a list of 11405 and 12257 phosphopeptides in relapsed and non-relapsed samples and aggressive and indolent cell lines, respectively. Furthermore, we conducted the ttest with permutations in order to find differently phosphorylated peptides between sample groups – i) early relapse and late relapse cases and ii) aggressive and indolent cell lines. From this analysis we found that 706 phosphopeptides were significantly differently phosphorylated when we compared relapsed vs. non-relapsed cases (FDR<0.15). Of this, 161 phosphopeptides were up-regulated (more phosphorylated) in relapsed vs non-relapsed cases and 545 phosphopeptides were down-regulated (less phosphorylated in relapsed vs. non-relpased cases). For cell lines, the same analysis yield in total 874 phosphopeptides that were significantly differently phosphorylated in aggressive vs. indolent cell lines (FDR<0.15) out of which 53 were found more phosphorylated and 821 were less phosphorylated in aggressive vs. indolent cell lines. The representative heat-map showing results of differential analysis of patient samples and cell lines is depictured in Figure 8 (A,B).

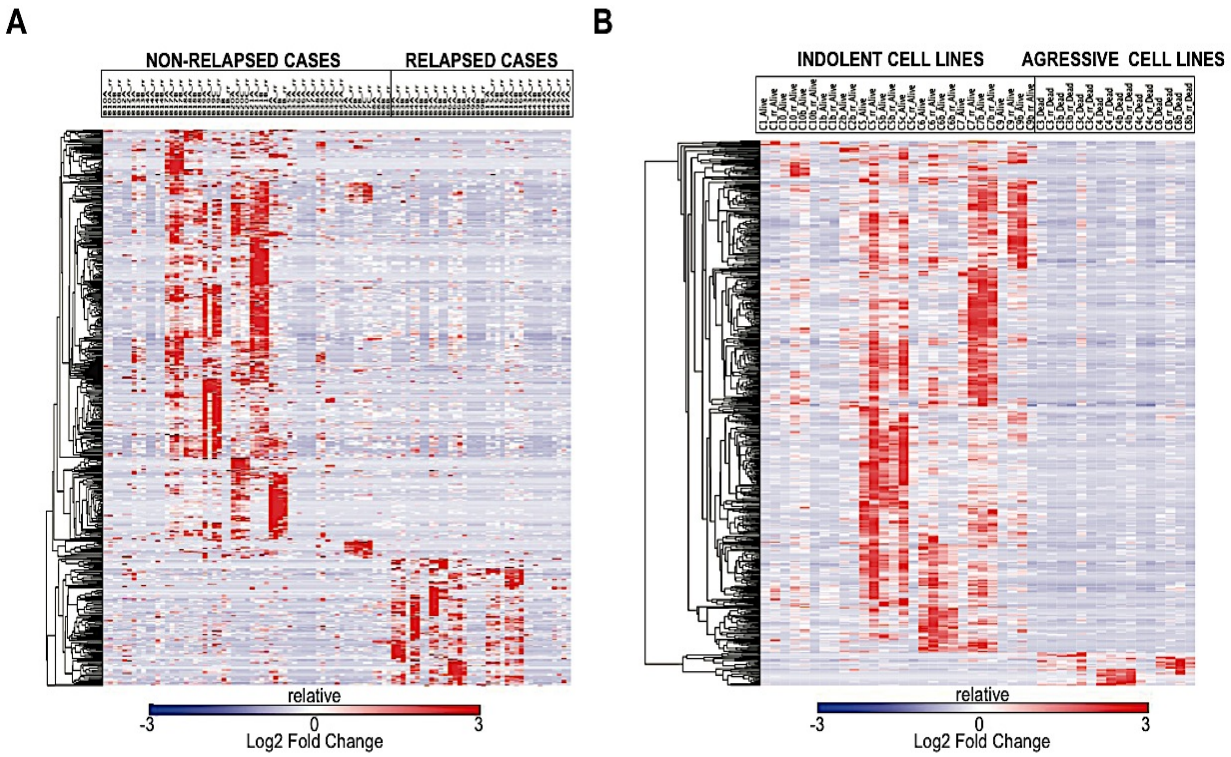


Figure 8. Heat-Map depicturing significantly differently phosphorylated peptides in tumors (by relapsing status) and cell lines (by aggressiveness groups), respectively. (A) Differential analysis of phospho-data of patient samples showed that 706 phosphopeptides were significantly differently phosphorylated in relapsed vs. non-relapsed cases. Out of 706, 161 were up-regulated (more phosphorylated) and 545 phosphopeptides were downregulated (less phosphorylated) in relapsed cases. (B) Differential analysis of phospho-data of cell lines showed that 874 phosphopeptides were significantly differently phosphorylated in aggressive vs indolent cell lines. Out of 874, 53 were up-regulated and 821 were downregulated in aggressive vs. indolent cell lines.

1.4. Kinase prediction and Enrichment analysis

Protein kinase target substrates typically depends on the primary amino acid sequence proximal to the site that will be modified (Dephoure,N. et all, Proc.Natl.Acad.Sci.U.S.A 2008). This sequence motif can be used for the identification of kinase specific substrates. In order to implement the phospho-information into further level of interpretation we performed motif-dependent categorization of the detected phosphopeptides, both in relapsed and non-relapsed cases and aggressive and indolent cell lines. Subsequently we applied a methodology similar to that used for finding Gene Set Enrichment (GSEAs)[121] among clusters of interest of gene expression data by applying annotation of kinase motifs extracted from Perseus software, which

integrates 327 linear motifs (80 of which are SH2-binding domain motifs, 23 are phosphatase substrate motifs and 224 kinase). The enrichment plots for the kinases driving the profile of relapsed versus non-relapsed case and aggressive versus indolent cell lines are presented in the Figure 9. Figure 9A shows one sample chart of P70S6K, whose increased function was found in the relapsed cases on the bases of the abundance of its putative identified substrates, shown in the right panel. The enriched kinases in the relapsed cases and aggressive cell lines are shown in Figures 9B and C, evidencing a further degree of overlap beyond GO pathways between tumor biopsies and cell lines. Interestingly besides the well-studied AKT kinase, CDK4/6[130], CDC2/CDK1[131] and PKC-Epsilon[132] that were to some extent previously linked to breast cancer biology, the reminder have not been proposed neither as main drivers of adverse clinical behaviour or as potential therapeutic target in TNBC.

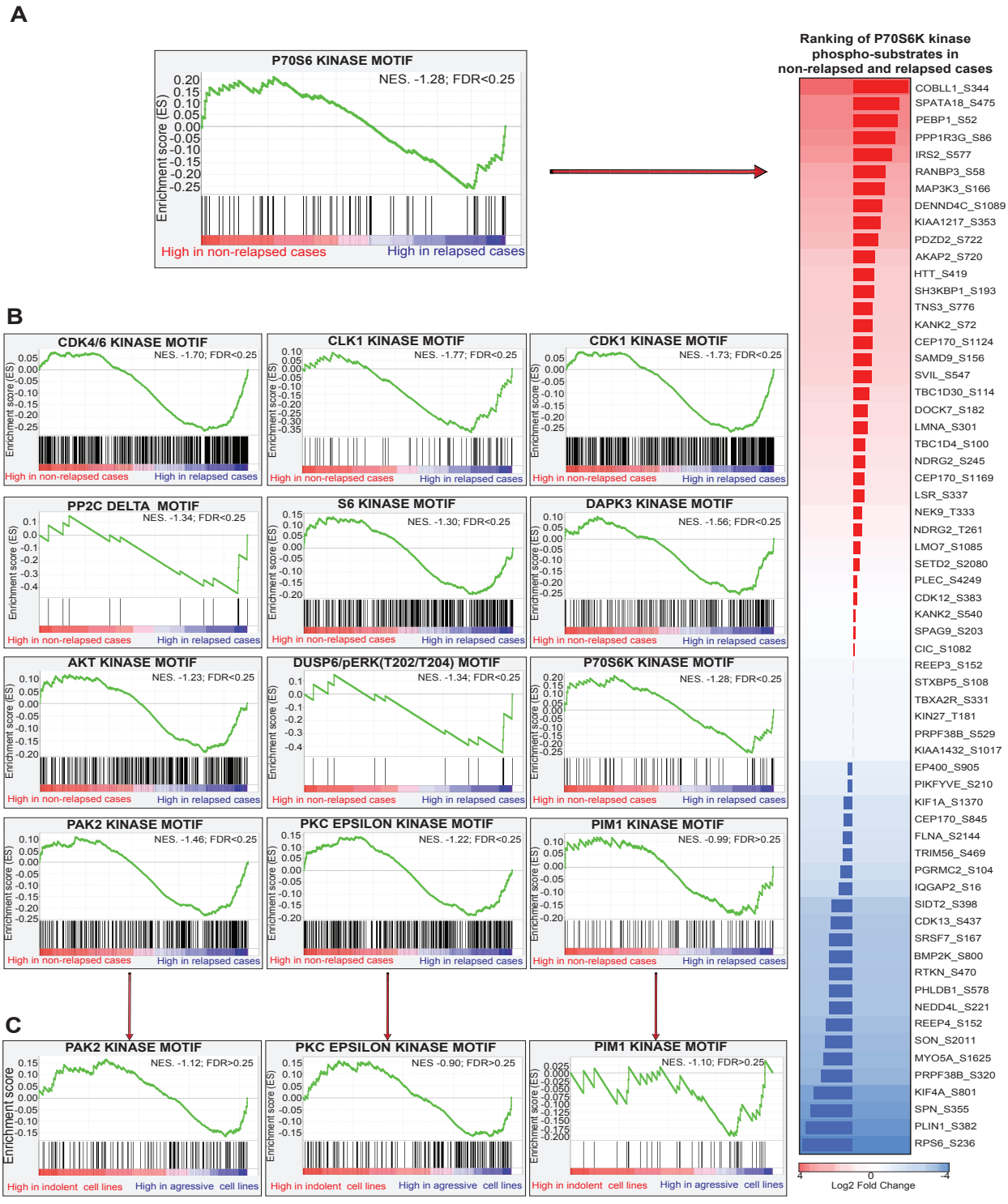


Figure 9. Kinase driving the profiles of the relapsed cases and aggressive TNBC cell lines. (A) Example of normalized enrichment score (NES) chart (left) for P70S6K obtained from the relative abundance of its phosphorylated substrates in relapsed and non-relapsed cases (right). The proteins are presented in the right column by their encoding gene adjacent to the site in which phosphorylation was detected. (B) The function of two phosphatases (DUSP6 and PP2C-delta) and ten kinases was found enriched in relapsed cases; three of which (C) were found as well in the aggressive cell lines.

2. VALIDATION AND TRANSLATION OF PHOSPHOPROTEOMIC SIGNATURE INTO A CLINICALLY ASSESSABLE TOOL

Differential analysis of phospho-proteomic data revealed that out of 11405 and 12257 phospho-sites identified by MS in TNBC patient samples and cell lines, 161 and 53 phospho-sites were significantly up-regulated (more phosphorylated) in relapsed cases and aggressive cell lines, respectively. Moreover, kinase prediction and subsequent GSEA analysis, pointed out main kinases that potentially drive aggressive clinical course of TNBC. In order to ascertain that the potential novel phospho-sites and driver kinases identified in training set are not simply the result of an imperfect *in silico* prediction we sought for external validation in an independent patient series and preclinical models.

2.1. Selection of antibodies for external validation

So far, it is not realistic to implement mass spectrometry into daily routine clinical practice mainly because of poor suitability of formalin fixed paraffin embedded samples (FFPE, the most widely available source of tumor material as a product of standard tissue processing procedures followed by surgical pathology laboratories) for the technique[133], time-to-storage-impact in the preservation of PTMs (port-translational modifications)[134] or the cost. For that reason we aimed to translate the information obtained with mass spectrometry to data that could be determined in FFPE. Translation of mass spectrometry into immunohistochemistry could involve two main challenges. First challenge is the reagent availability, meaning that the antibody suitable for IHC-p is not always available for all detected kinases or upregulated phosphosites. Second, it is important to determine how we define what is an “activated kinase” and moreover how to detect in FFPE its active conformation. In order to select the antibodies that could be used to detect, in a representative way, the activated form of a kinase under study, we followed certain steps depictured in Figure10.

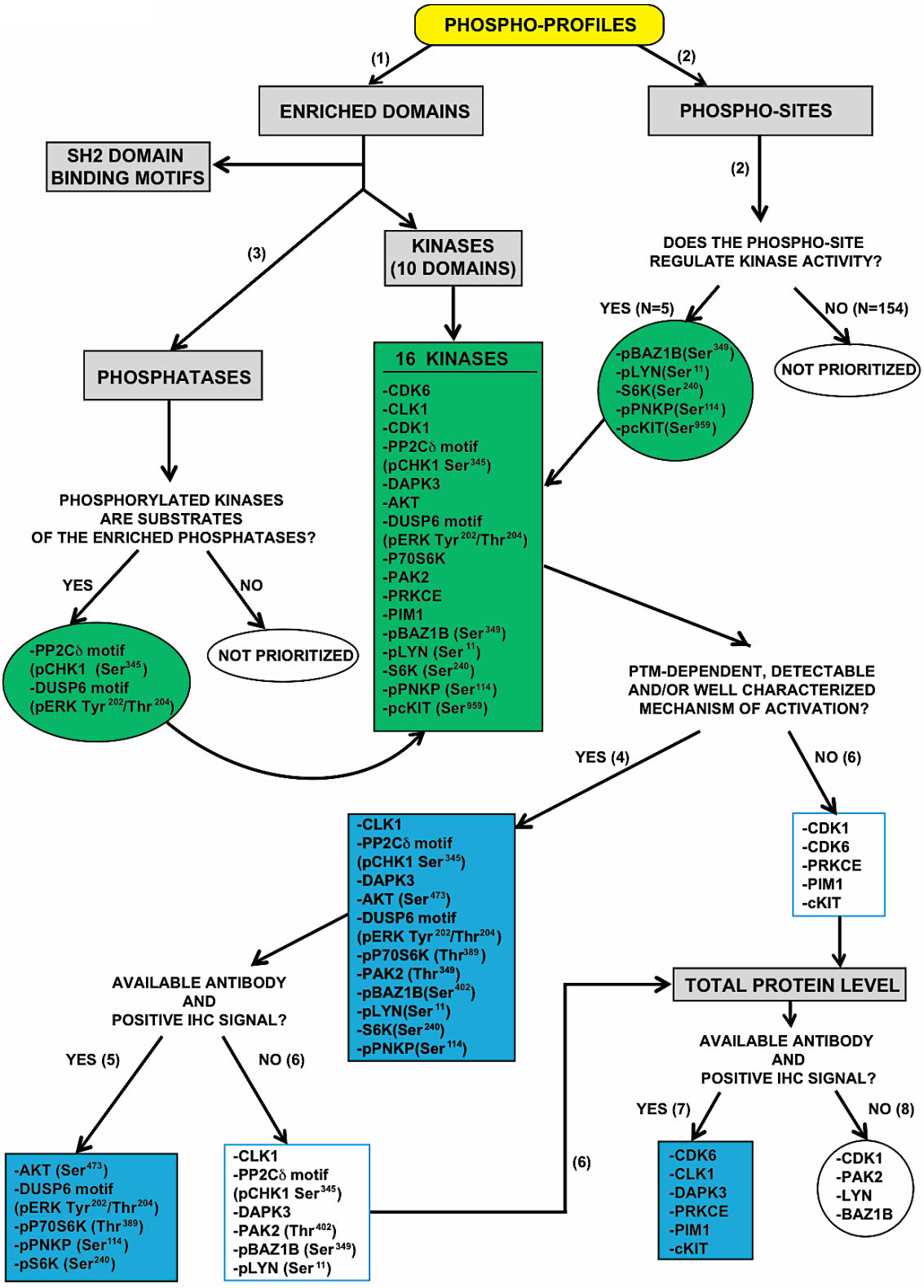


Figure 10. Workflow for antibody selection.

In the first step (1) we prioritized enriched-domains over phospho-sites, since they integrate a broader portrait of the effector layer from the functional point of view. Additionally, we aimed to find classifiers that are at the same time targetable enzymes and for that reason the non-kinase hits we did not considered for the validation. Nevertheless some of the phospho-sites (2) that were found up-regulated in relapsed cases correspond as well to kinases and kinases are very often regulated by posttranslational modifications like phosphorylation. Thus, we added five kinases to testing workflow. However some of the enriched-domains found, corresponded to phosphatases (3) (DUSP6 and PP2C delta), which reflect the abundance of the phosphorylated substrate that is cleaved by the predicted phosphatase. If one of those substrates was a phosphorylated form of a kinase than this hit is added to the testing workflow as well. In total sixteen kinases were thus selected for the validation step. In cases a specific kinase had well-described, dominant and theoretically detectable mechanism of activation (4) (Ser 473 in the case of AKT[135]), antibodies against that targets were tested first in the pellet of tnbc cell lines embedded in paraffin with known status of targets by western blot (Figure 11). Those that yielded IHC signal were tested in validation set. Following this workflow five hits were selected (5). For those that failed detection or lack the antibody suitable for IHC in human tissue, we tested non-phosphorylated forms of the kinases (6). Additionally, three more types of kinases were probed with antibodies against the total proteins (6). These include, the kinases that do not need any further posttranslational modification for its activation (i.e. PIM[136]), the kinases which activation status is not detectable by IHC (i.e., PRKCE[137,138] or KIT[139,140]) and finally the kinases which activation is driven by complex and transient mechanism with even opposing action (PRCKE[141] or KIT[142,143]). We used the same validation set-up workflow which led to another six probes that were suitable for detection (7) and four probes that were discarded.

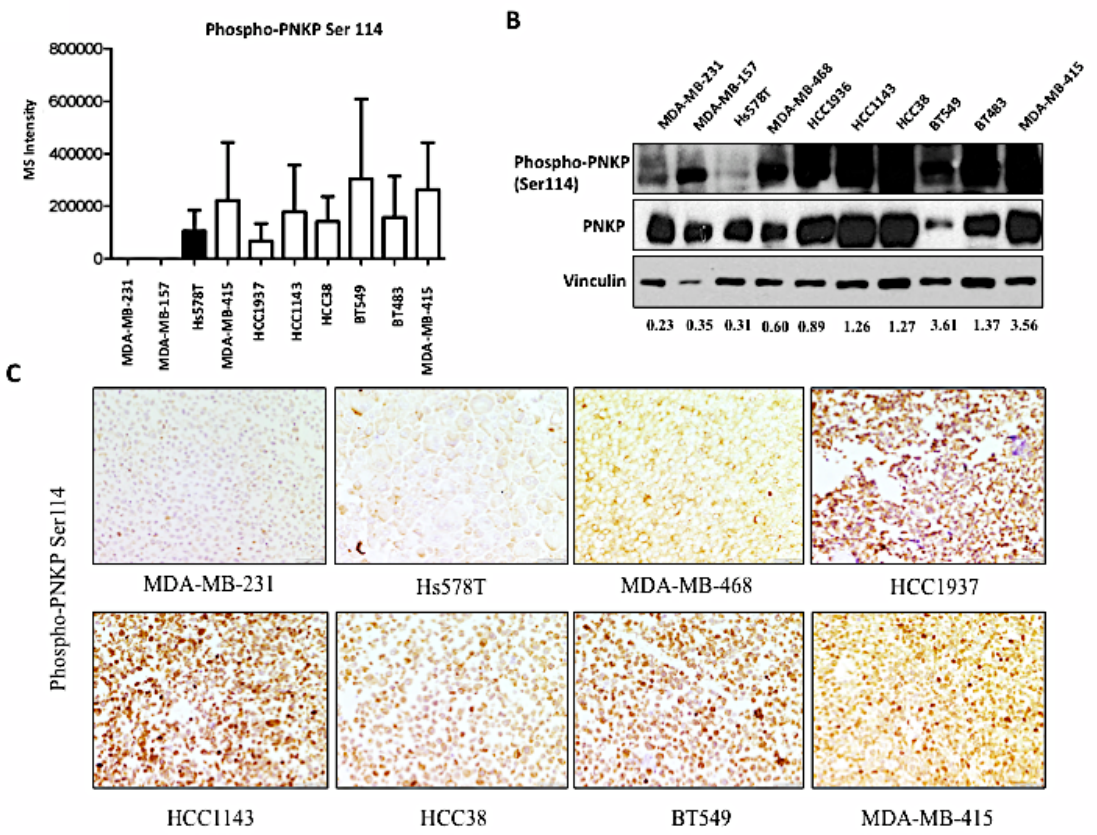


Figure 11. Validation of phospho-PNKP Ser114 in TNBC cell lines (A) Graphs showing average intensities of phospho-PNKP Ser114 quantified by mass spectrometry in TNBC cell lines. Mean and StDev were calculated from minimum 12 replicates of MS injections.(B) Western blot analysis of phospho-PNKP Ser114 in TNBC cell lines. (C) IHC-p with antibody against phospho-PNKP Ser114 in TMAs from TNBC cell lines.

The workflow we used (depictures in Figure 10 and Figure 11) yielded 11 valid antibodies against the following targets: CDK6, CLK1, pS6K (Ser240), DAPK3, PRKCE, pAKT (Ser473), pERK (Thr202/Tyr204), pP70S6K (Thr389), PIM1, pPNKP (Ser114/Thr118) and cK

2.2. Validation of selected phospho-sites and kinases in validation set

In order to validate selected targets we gathered an independent validation set of 113 consecutive TNBC patient samples embedded in paraffin from which tissue microarrays (TMAs) were made. Table 2 and Figure 12 shows clinical characteristics and survival curves of 113 patients from validation set

	All (N=113)	Relapsed (N=41; 37%)	Non-relapsed (N=72, 63%)
Age (mean, range)	56.7 (26-78)	59.3 (28-78)	55.0 (26-77)
TNM Stage			
I	26 (23%)	3 (10%)	23 (32%)
II	63 (55%)	25 (60%)	38 (52%)
III	24 (22%)	13 (30%)	11 (16%)
Type of chemotherapy			
None	12 (10%)	8 (19%)	4 (5%)
CMF only	33 (29%)	9 (22%)	24 (33%)
Anthracyclines plus cyclophosphamide	41 (36%)	11 (27%)	30 (42%)
Antracyclines plus cyclophosphamide plus taxanes	27 (25%)	13 (32%)	14 (20%)
Ki67 average (range)	39.8% (1%-97%)	39.7% (1%-97%)	39.9% (1%-80%)
Grade			
G1	4 (3%)	1 (2%)	3 (4%)
G2	31 (27%)	13 (32%)	18 (25%)
G3	78 (70%)	27 (66%)	51 (71%)
Premenopausal	43 (38%)	11 (27%)	32 (44%)

Table 2. Clinical and demographic characteristics of the validation set

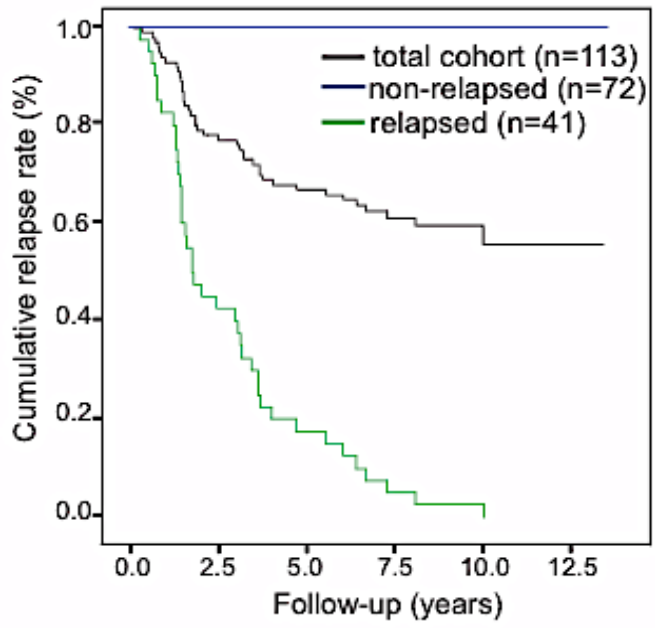


Figure 12. Validation set. Survival curve of the 113 patients from the independent validation set (black); 72 of these 113 patients did not relapse in more that 12.5 years of observation (blue) whereas 41 patients relapsed (green).

Next, we performed immunohistochemistry (IHC) on TMAs with 11 valid antibodies selected on the bases of workflow shown in Figure10. To acquire and quantify the IHC signal we used automated Leica microscope and Ariol SL-50 software. For each biopsy, IHC score was generated and presented as a continuous variable from 0-3. In Table 3 and Figure 13, the IHC score distribution (in quartiles) and representative staining examples are shown.

	Upper Quartile(75)	Medium Quartile(50)	Lower Quartile(25)
CDK6	0.0783	0.0280	0.0073
pPNKP Ser114/Thr118	0.6968	0.4562	0.2464
cKIT	0.0057	0.0029	0.0011
PRKCE	0.1754	0.0515	0.0170
pERK Thr202/Tyr204	0.4075	0.1099	0.0168
pP70S6K Thr389	0.2728	0.1383	0.0295
pAKT Ser473	0.2206	0.1084	0.0365
CLK1	1.0049	0.7888	0.6007
pS6K Ser240/Ser244	0.2395	0.1048	0.0406
DAPK3	0.9532	0.7504	0.5642
PIM1	0.3742	0.1815	0.0735

Table 3. IHC-scores divided in quartiles for each tested probe.

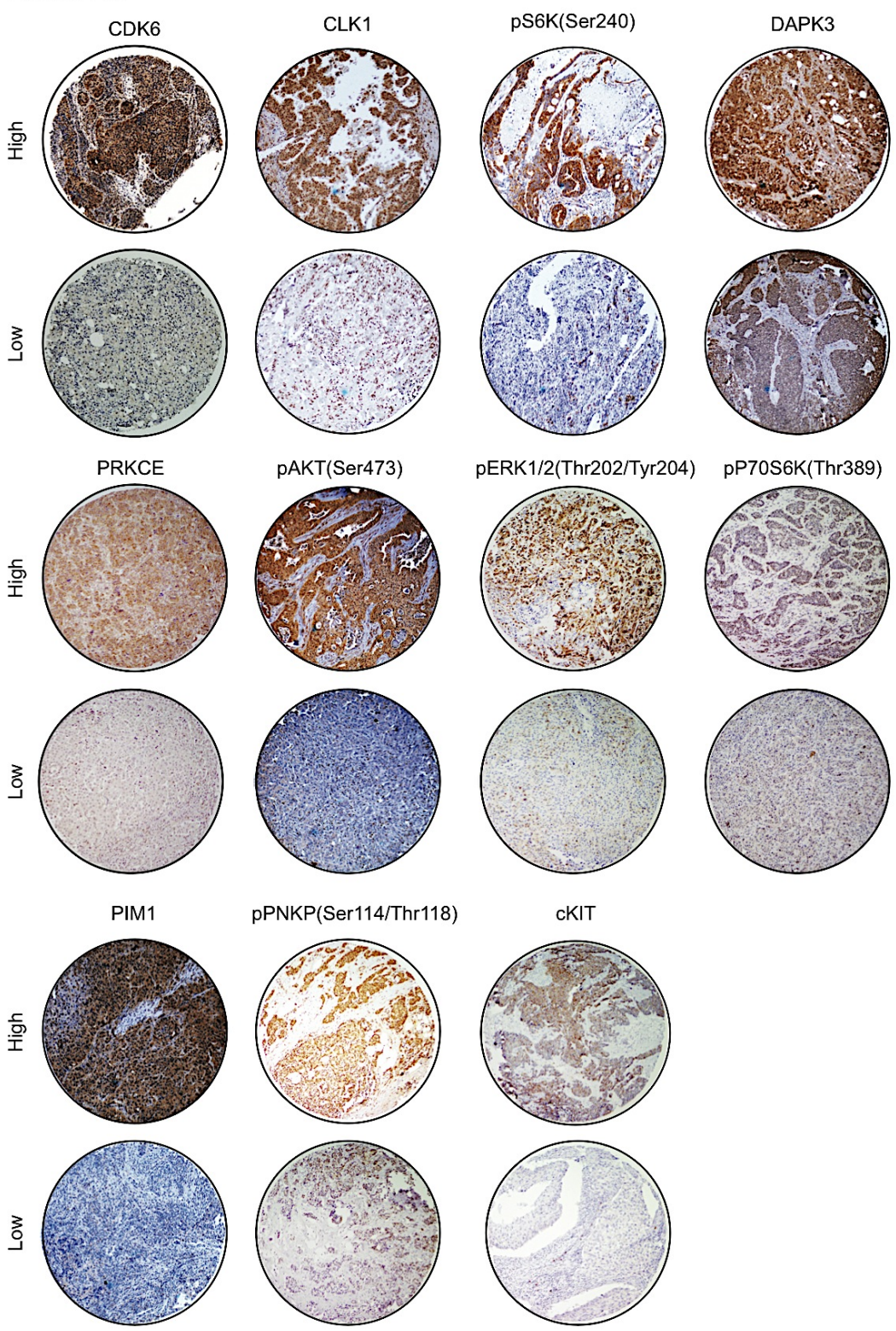


Figure 13. Representative staining imagines of upper/lower quartiles for each tested probe.

After obtaining IHC scores for all biopsis, in order to evaluate if validated targets show prognostic power, for each kinase we compared time to relapse between patients with high activity (the IHC score in upper quartile) versus the remainder. Out of eleven kinases that were predicted to have high activity in the relapsed cases of the training set, six of them had the ability to discriminate the clinical course in the validation set. This is shown in Figure 14A by the Kaplan-Meier curves and hazard rations adjusted by T, N, grade and age for patients that had IHC scores in upper quartile versus the reminder. The other five kinases did not show significant prognostic value (Figure 14B).

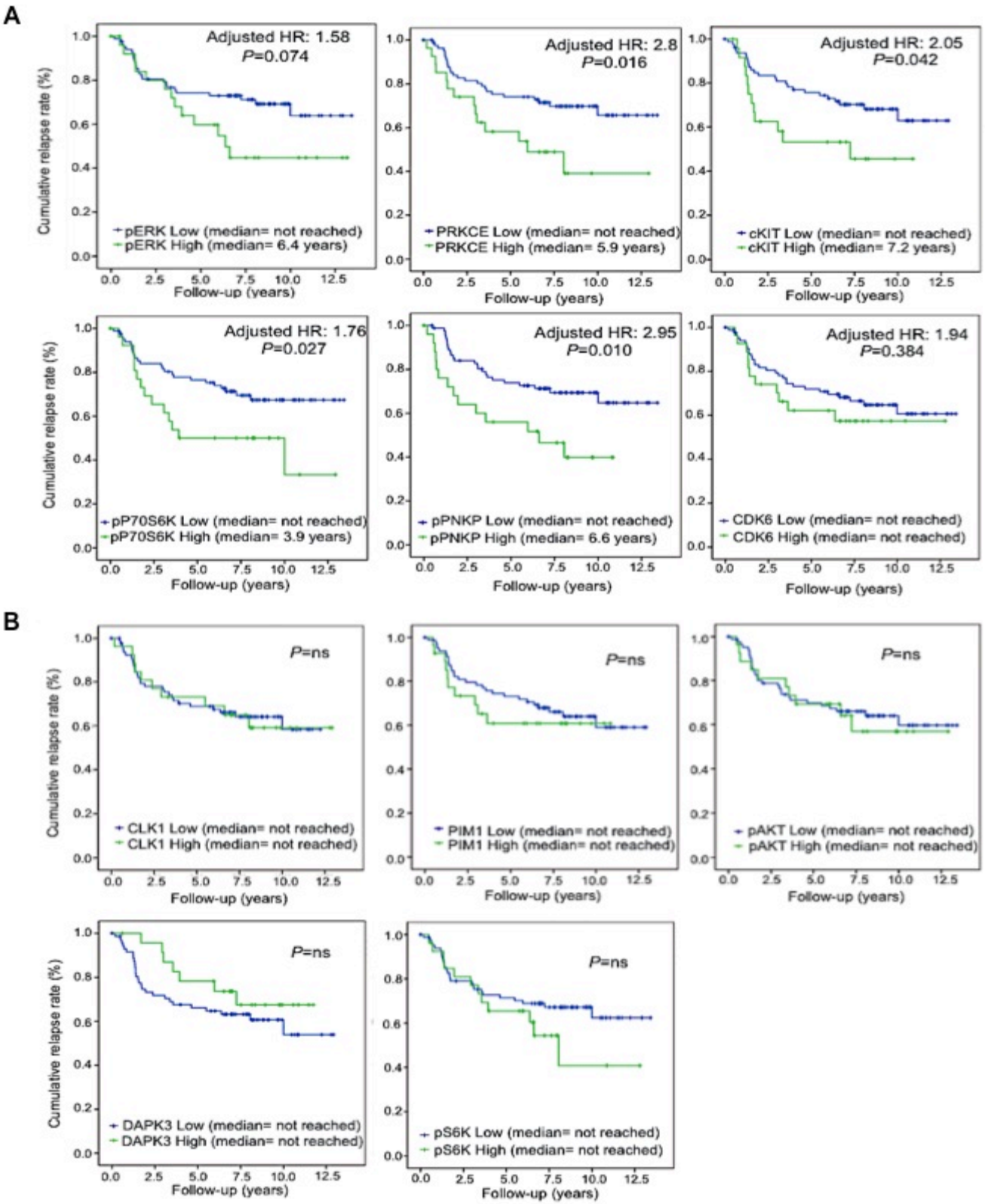


Figure 14. Kaplan-Meier curves of 113 patients divided based on presents (green) or absents (blue) of upper quartile staining of tested probes. (A) Represents Kaplan-Meier curves of pP70S6K (Thr389), PRKCE, c-KIT, pERK (Thr202/Tyr204), pPNKP (Ser114/Thr118) and CDK6, probes which upper quartile

staining was able to discriminate those patients that are relapsing from the others.(B) Kaplan-Meier curves of CLK1, PIM1,pAKT (Ser473), DAPK3 and pS6K (Ser240) which upper quartile staining did not pass statistical significant. The shown P values correspond to the Log-rank test for the median overall survival comparison. The depicted hazard ratio correspond to the adjusted hazard ratio in a Cox model by T, N, G and age (P<0.05).

2.3. Kinomic landscape of triple negative breast cancer (TNBC): clinical implications

We identified six kinases which high activity seem to be associated with adverse clinical outcome. Subsequently we wondered whether the effect of each of the six kinases with prognostic value was independent from the effect of the other kinases. With this aim, we performed linear regression analysis. As shown in Figure 15, we observed lack of linear relationship between kinases under comparison.

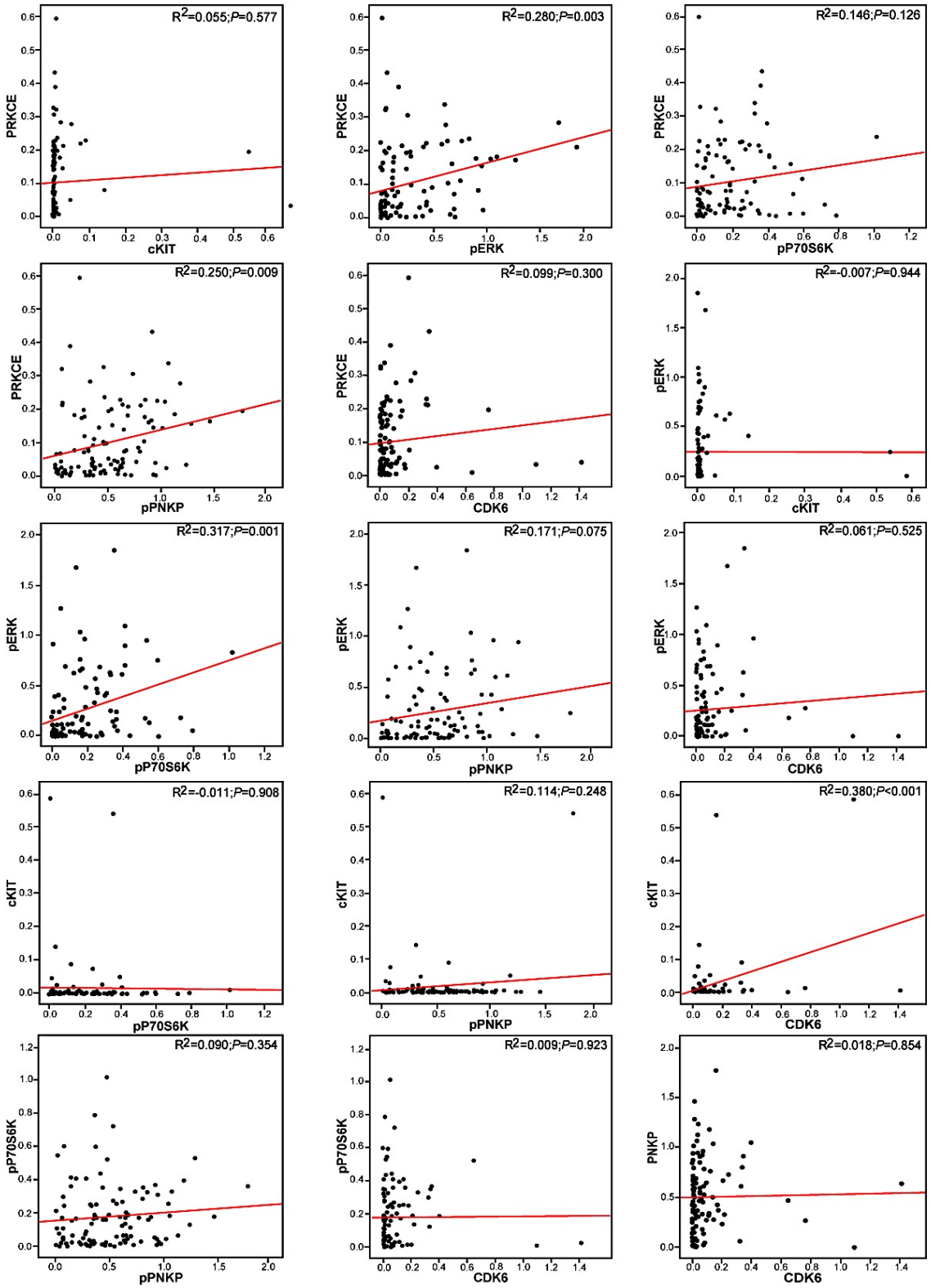


Figure 15. Lack of significant co-linearity between six kinases with independent prognostic value. We performed linear regression analysis of 15 possible combinations of the 6 kinases with prognostic impact. 11 of them showed lack of statistical significance. The reminder, although significant, had a very low R2-coefficient.

Moreover in order to see the distribution of different staining patterns of kinases with upper-quartile IHC-score, we sorted cases from validation set based on the relapsing status and presents or absents of each of six kinases. Landscape of activation of six kinases in cases from validation set is depictured in Figure 16.

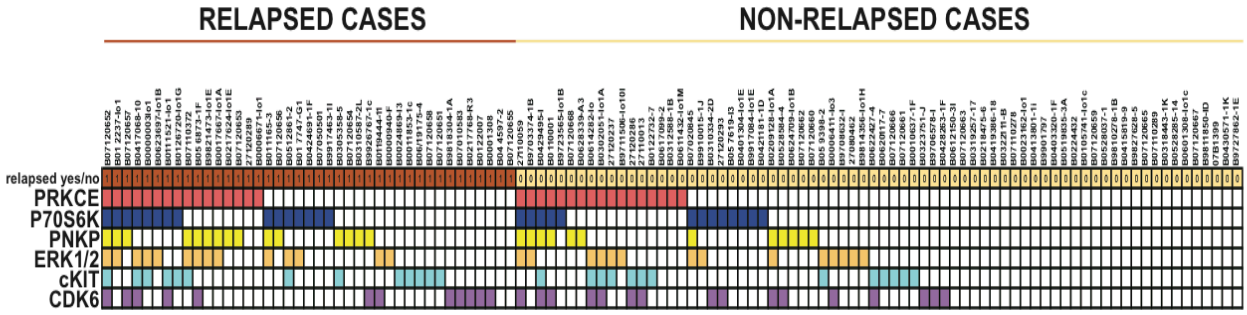


Figure 16. The kinome landscape of TNBC cases. Samples from validation set are ordered horizontally based on the status of relapse. Each of the six kinases with prognostic power are listed on the left had side of the grid. The presents of activated kinase is highlighted with different color, PRKCE (red), P70S6K (dark blue), PNKP (yellow), ERK1/2 (orange), cKIT (light blue) and CDK6 (purple) so in this way the number of activated kinases per patient is depicted.

After obtaining the grid we could observe clear aggregation of activated kinases in the relapsed cases with several different staining patterns of kinases with upper-quartile staining IHC scores. Interestingly we found only two patients in the relapsing group that did not have kinases in upper-quartile IHC scores. Ability to identify patients who, after standard chemotherapy treatment, could be effectively treated compared with patients who might need different therapeutic regiment relapse seams indeed tempting. Thus we examine the outcome of patients whose tumors show high activity of any of six kinases versus the remainder. The resulting combined variable (herein “K-high”) allowed division of the 113 patients in two groups. Former with 81 patients (71%, which tumors show high activity of any of six kinases) and latter with 32 patients (29%, which tumors show low activity of six kinases). The relapse rate were 47% and 7%, respectively, after more that 12 years of follow-up. Respective median time to relapse were 9.9 years and not reached (P<0.001) (Figure 17). After adjusting for T, N, G and age, K-high showed a hazard ratio for relapse of 9.22 (P=0.002).

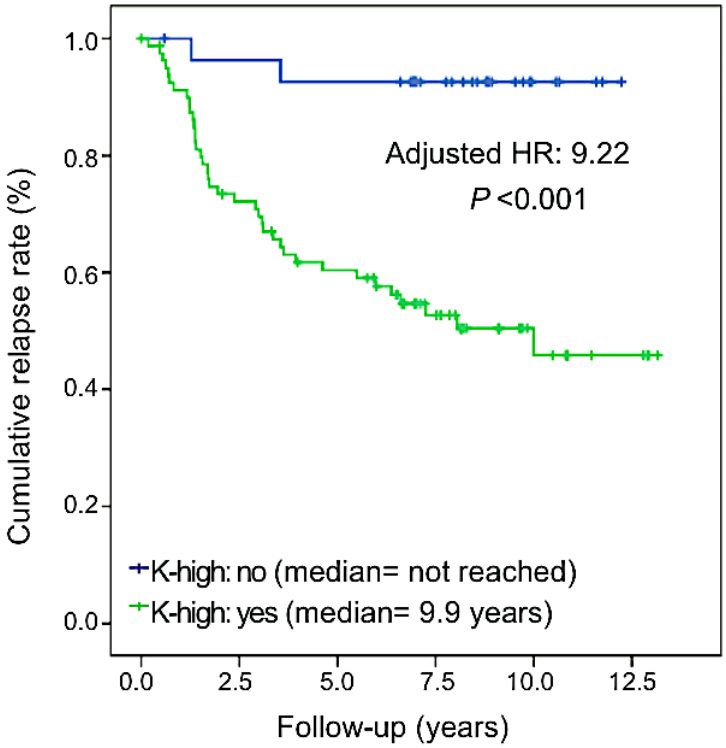


Figure 17. Kaplan-Meier curve for combined variable ("K-high"). Relapse of patients whose tumors show high activity of any of 6 kinases (green; K-high=yes) versus the rest of the patients from validation set (blue; K-high=no).

Interestingly the number of activated patterns per patient, as it could be observed in the grid (Figure16), shows important variation. Out of 64 possible patterns we found only 38 listed in Table 4. Twelve of them were exclusive for relapsed cases. However majority were present in very few cases. The most observed pattern was lack of activation of any of six kinases and this pattern was present in most of cured cases (29 patients). The other very common pattern was activation of single kinases (c-KIT or CDK6). Nevertheless great number of patterns involved combination of two or more activated kinases.

PATTERN NUMBER	WHAT KINASES ARE IN THE PATTERN	HOW MANY PATIENTS HAVE THE PATTERN	HOW MANY CURED (0) CASES, HOW MANY RELAPSED (1) CASES
1	PRKCE; P70S6K; PNKP; ERK1/2; cKIT; CDK6	1	1 (1)
2	PRKCE; P70S6K; PNKP; ERK1/2	2	1 (1); 1 (0)
3	PRKCE; P70S6K; PNKP; CDK6	2	1 (1); 1 (0)
4	PRKCE; P70S6K; PNKP; ERK1/2; CDK6	1	1 (0)
5	PRKCE;P70S6K;PNKP;cKIT;CDK6	1	1 (0)
6	PRKCE; P70S6K; ERK1/2; cKIT; CDK6	1	1 (1)
7	PRKCE; P70S6K; ERK1/2; cKIT	1	1 (1)
8	PRKCE; P70S6K; ERK1/2	1	1 (1)
9	PRKCE; P70S6K; cKIT; CDK6	1	1 (1)
10	PRKCE; P70S6K; cKIT	1	1 (1)
11	PRKCE; P70S6K	1	1 (0)
12	PRKCE; PNKP; ERK1/2; cKIT	1	1 (1)
13	PRKCE; PNKP; ERK1/2; CDK6	1	1 (1)
14	PRKCE; PNKP; ERK1/2	2	2 (1)
15	PRKCE; PNKP	4	2 (1); 2 (0)
16	PRKCE; ERK1/2; cKIT; CDK6	2	2 (0)
17	PRKCE; ERK1/2; cKIT	1	1 (0)
18	PRKCE; ERK1/2	1	1 (0)
19	PRKCE; cKIT; CDK6	2	2 (0)
20	PRKCE; cKIT	1	1 (0)
21	PRKCE	5	2 (1); 3 (0)
22	P70S6K; PNKP; ERK1/2	2	1 (1);1 (0)
23	P70S6K; PNKP	1	1 (1)
24	P70S6K; ERK1/2; cKIT	1	1 (1)
25	P70S6K; ERK1/2	2	1 (1); 1 (0)
26	P70S6K;CDK6	2	2 (0)
27	P70S6K	7	3(1); 4 (0)
28	PNKP; ERK1/2; CDK6	1	1 (0)
29	PNKP; cKIT	1	1 (1)
30	PNKP; CDK6	2	1 (1); 1 (0)
31	PNKP	5	2 (1); 3 (0)
32	ERK1/2; cKIT	1	1 (0)
33	ERK1/2; CDK6	2	1 (1); 1 (0)
34	ERK1/2	4	1 (1); 3 (0)
35	cKIT; CDK6	1	1 (0)
36	cKIT	9	5 (1); 4 (0)
37	CDK6	8	5 (1); 3 (0)
38	non of kinases activated	31	2 (1); 29 (0)
		113	

Table 4. Kinase activation patterns found in the validation set. For each pattern the description of its components (activated kinases), the total number of patients and number of cases with each patter that are cured (0) or relapsed (1) is depicted in the table.

2.4. In vitro validation of kinases defining the TNBC landscape

By mass spectrometry we identified and quantified peptides that are phosphorylated as a result of the hyperactivation of the six kinases. Most likely this hyperactivation of the kinases is a consequence of a genetic evolution process, and for that reason kinases might be used as *de facto* targetable drivers. In order to test

weather any of six kinases could be used as targets for triple-negative breast cancer we conducted in vitro experiment in 10 triple-negative breast cancer cell lines. Firstly we estimated IC50 of drugs against each of the six kinases. We purchased A12B4C3, palbociclib, GDC-0994, imatinib, the peptide-EAVSLKPT-OH and Ly2584702 against PNKP, CDK6, ERK, c-KIT, PRKCE and P70S6K, respectively. As shown in Figure 18A and B, most of the single-agent treatments in each of the ten triple negative cell lines were in micromolar range and showed mild growth inhibition effect.

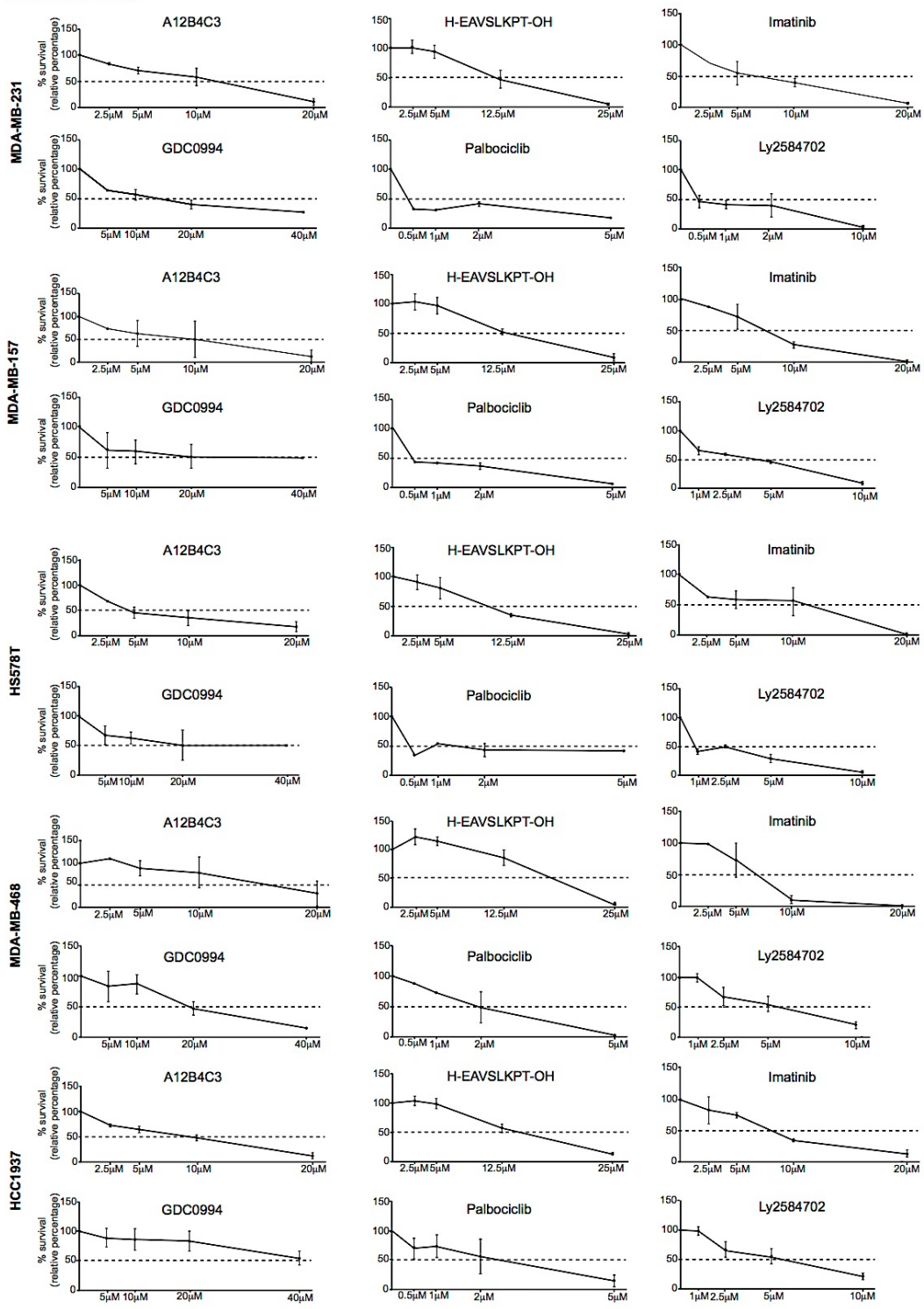


Figure 18. (A) Efficacy of single agents targeting the signaling nodes in K-high. Cell survival of ten triple-negative cell lines after treatment with increasing concentrations of six against (A12B4C3, H-EAVSLKPT-OH, Imatinib, GDC-0994, Palbociclib and Ly2584702).

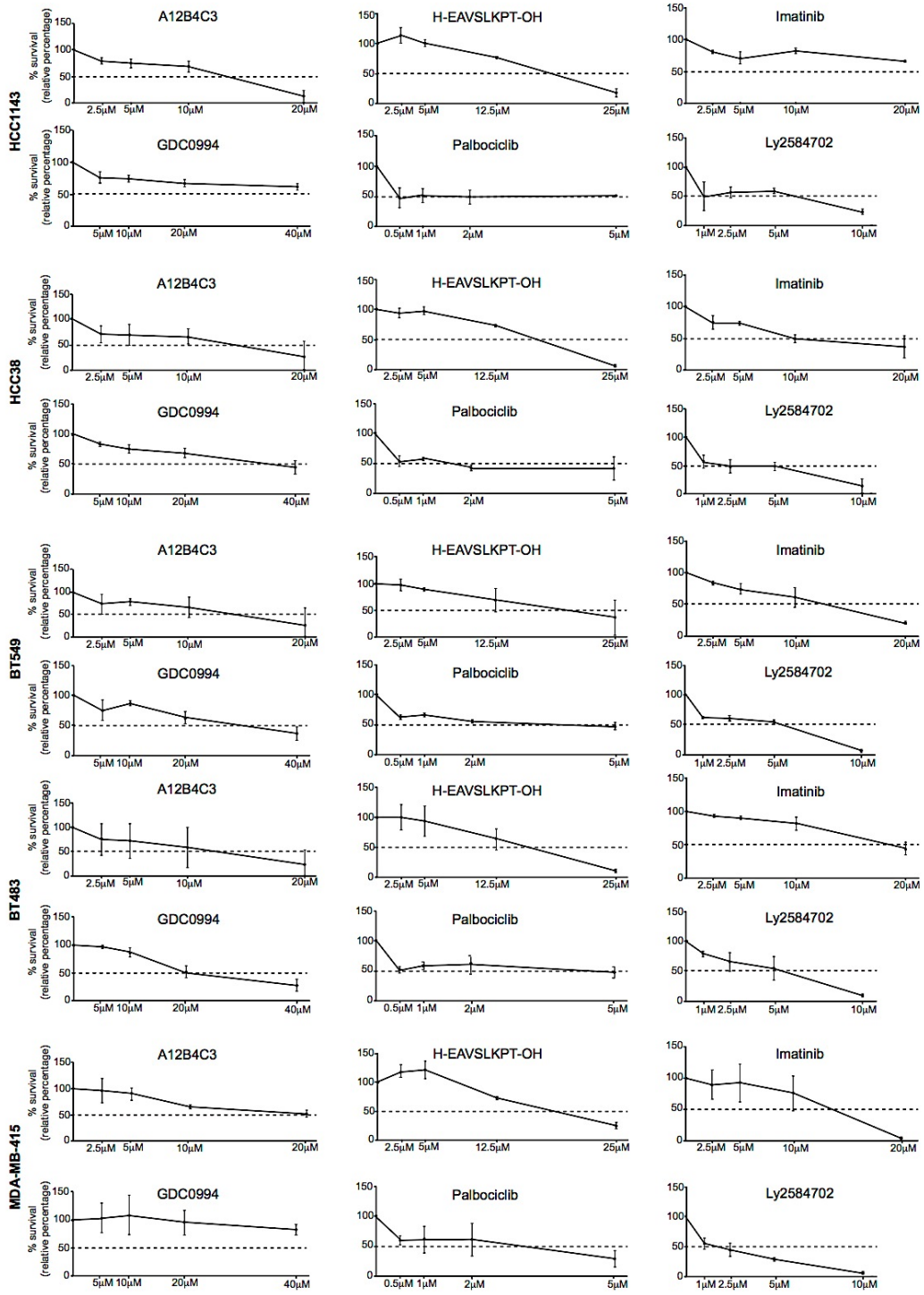


Figure 18. (B) Efficacy of single agents targeting the signaling nodes in K-high. Cell survival of ten triple-negative cell lines after treatment with increasing concentrations of six against (A12B4C3, H-EAVSLKPT-OH, Imatinib, GDC-0994, Palbociclib and Ly2584702).

However the lack of co-linearity between activated kinases (Figure15) and the observed trend toward aggregation of activated kinases in the relapsed cases (Figure16), made as thinking that probably additional effect of more than one kinase is necessary to maintain the behavior of the relapsing phenotype. For that reason we designed 2-by-2 drug combination pairs (15 pairs in total) to test in ten triple negative cell lines. Among 150 test combinations, 0 (0%) were antagonistic, 1 (0.7%) was additive, and 149 (99.3%) were supra-additive (Figure 19A and B). MDA-MB-468 cell lines treated with combination of PNKP inhibitor and PRKCE inhibitor peptide was the only combination showing the effect within additive boundaries.

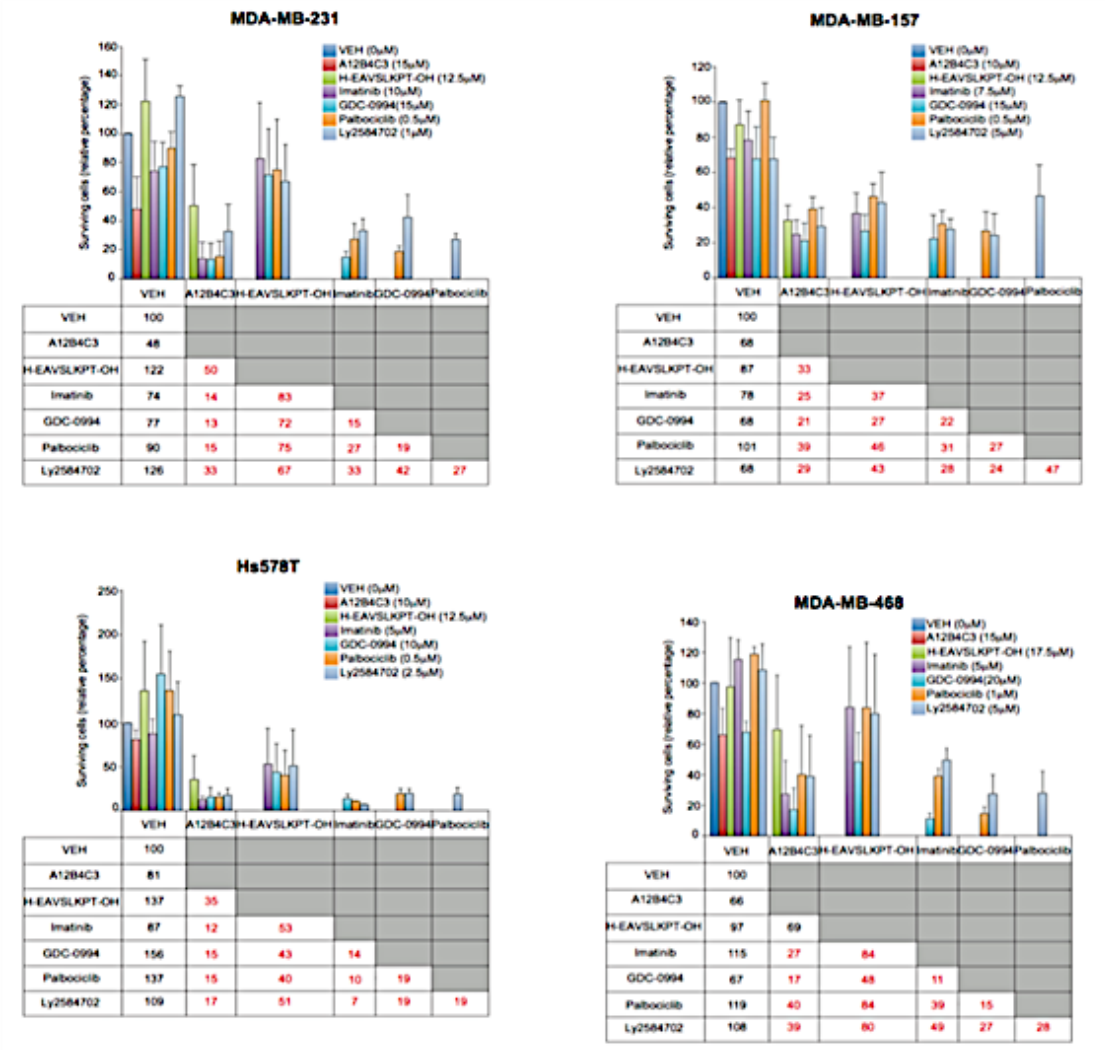


Figure 19. (A) Therapeutic synergism of inhibiting the driving kinases of TNBC in 4 of 10 triple-negative cell lines. Relative cell viability of each of 4 triple-negative cell lines after 72h of exposure to each of the compounds alone or the 15 possible combinations compared to vehicle alone (100%). Drugs were

applied at the concentrations as indicated in the figures. The percentages presented in red highlight if the value is beyond the limits of the additive effect.

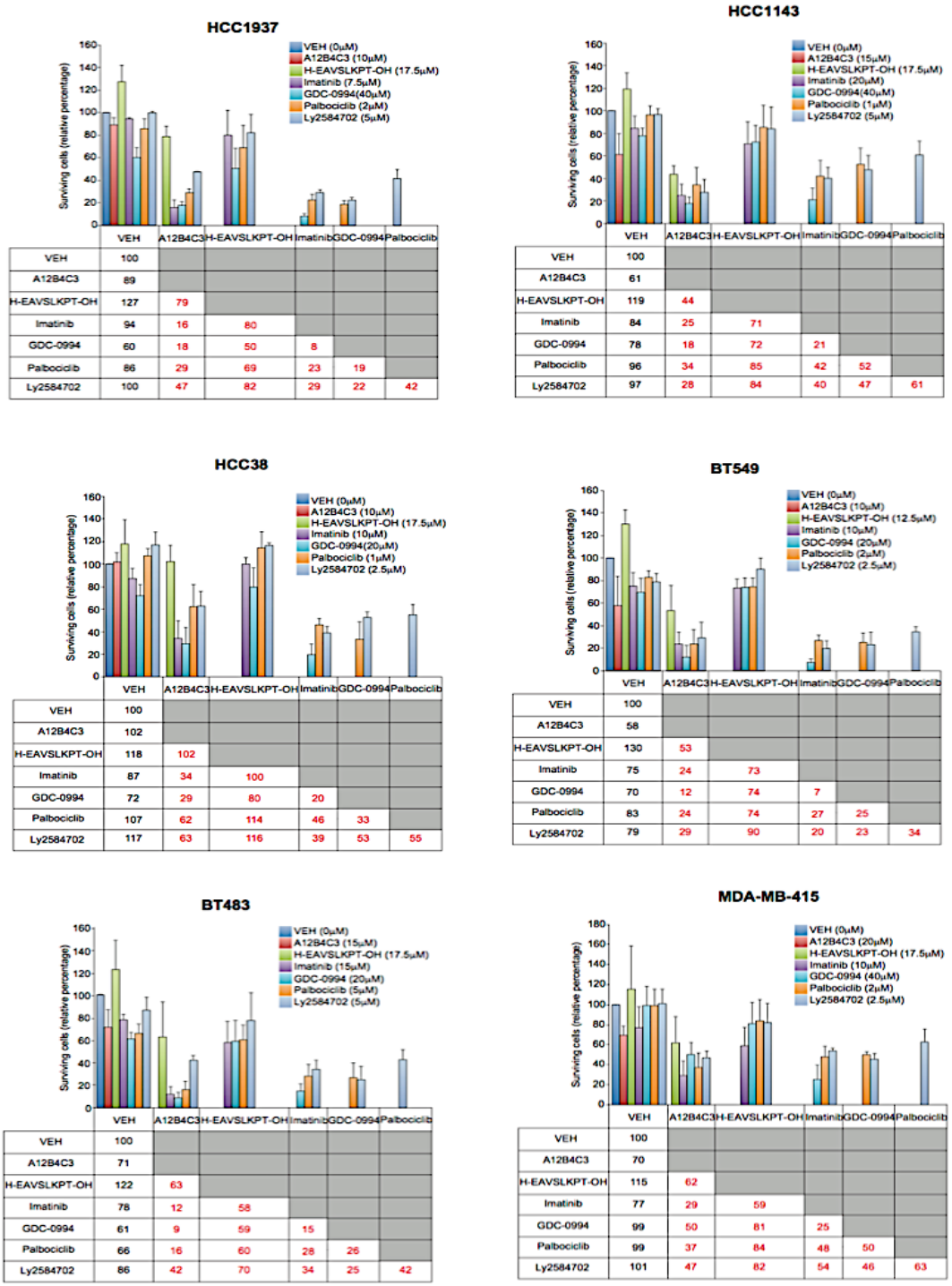


Figure 19. (B) Therapeutic synergism of inhibiting the driving kinases of TNBC in 6 of 10 triple-negative cell lines. Relative cell viability of each of 6 triple-negative cell lines after 72h of exposure to each of the

compounds alone or the 15 possible combinations compared to vehicle alone (100%). Drugs were applied at the concentrations as indicated in the figures. The percentages presented in red highlight if the value is beyond the limits of the additive effect.

2.5. Validation of effective drug combinations in vivo

After showed supra-additive effects of drug combinations in cell lines we next proceeded to validate this further in *in vivo* animal model. We selected one of the aggressive cell lines MDA-MB-231 for orthotopic xenograft model on the bases of expression of target proteins (Figure 20).

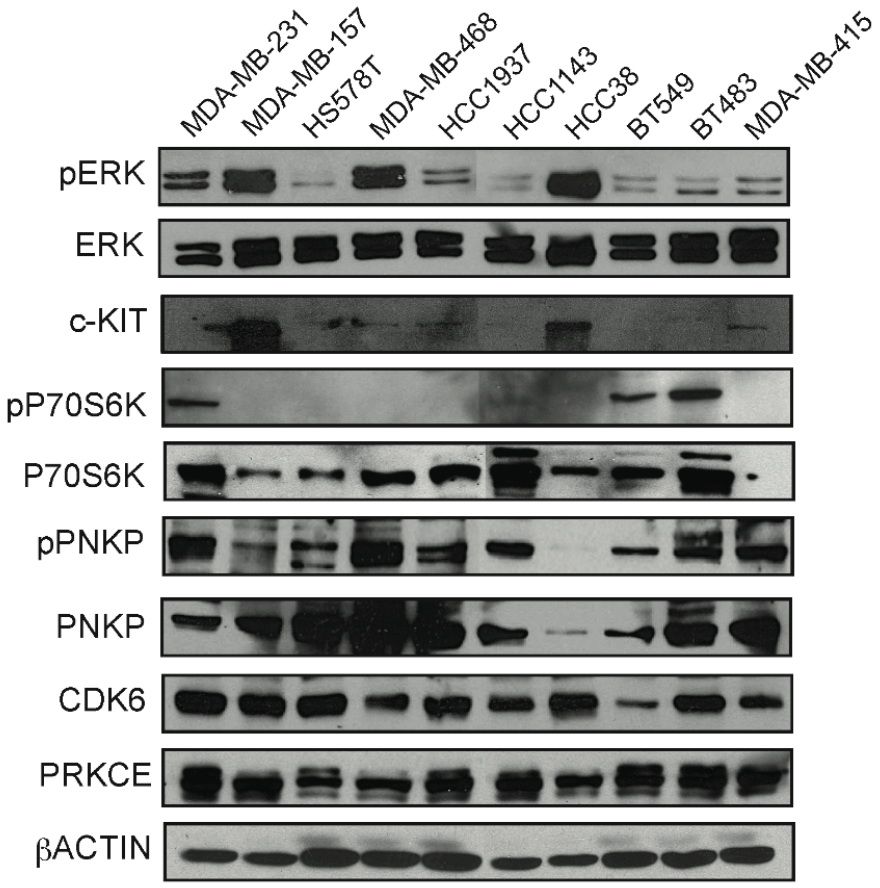


Figure 20. Therapeutic targets - levels of the 6 kinases of the signature in the cell lines. Immunoblots of lysates from TNBC cell lines. β-actin was used as loading control.

For orthotropic mouse model, tumors were allowed to grow proximally up to 500 mm^3 and then the animals were randomized to receive single agent or combination of two regiments (n>6 tumors per group). Since specific inhibitor for PNKP has not been tested yet for animal experiments, to avoid off-target effects and unknown pharmacodynamics and pharmacokinetics we decided to use CRISPR–Cas9 technology with small guiding RNA targeting PNKP (sgPNKP). We first verified efficient down- regulation of PNKP protein expression in MDA-MB-231 cell lines (Figure 21).

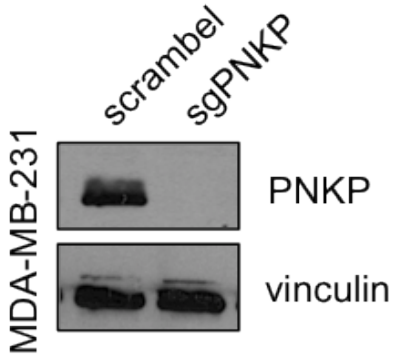


Figure 21. Immunoblotting of cell lysate from MDA-MB-231 cell line introduced with CRISPR construct.

Next the cells containing sgPNKP were injected into mammary fat pad and allowed to grow up to 500mm^3. Then the animals were randomized to receive vehicle or indicated drug combinations. Figure 22 A and B shows the tumor growth inhibition and tumor burden that we followed for 2 weeks. Our in vivo data indicate that single agent treatment show modest effects on tumor growth. The tumor growth inhibition was 35.5 with ERK inhibitor (GDC-0994), 20.8 with cKIT inhibitor (imatinib), 39.6 with CDK6 inhibitor (palbociclib) and 25.6 in case of tumors with downregulation of PNKP (sgPNKP). However combinatory treatments show significant decrease of tumor growth compared to single treatments. Tumor growth inhibition was 70.2 (imatinib plus GDC-0994), 90.0 (palbociclib plus GDC-0994), 55.3 (sgPNKP+imatinib) and 56.8 (sgPNKP+GDC-0994) at T=2 weeks. This inhibitory effect is further highlighted when considering tumor burden depicted in Figure 22B.

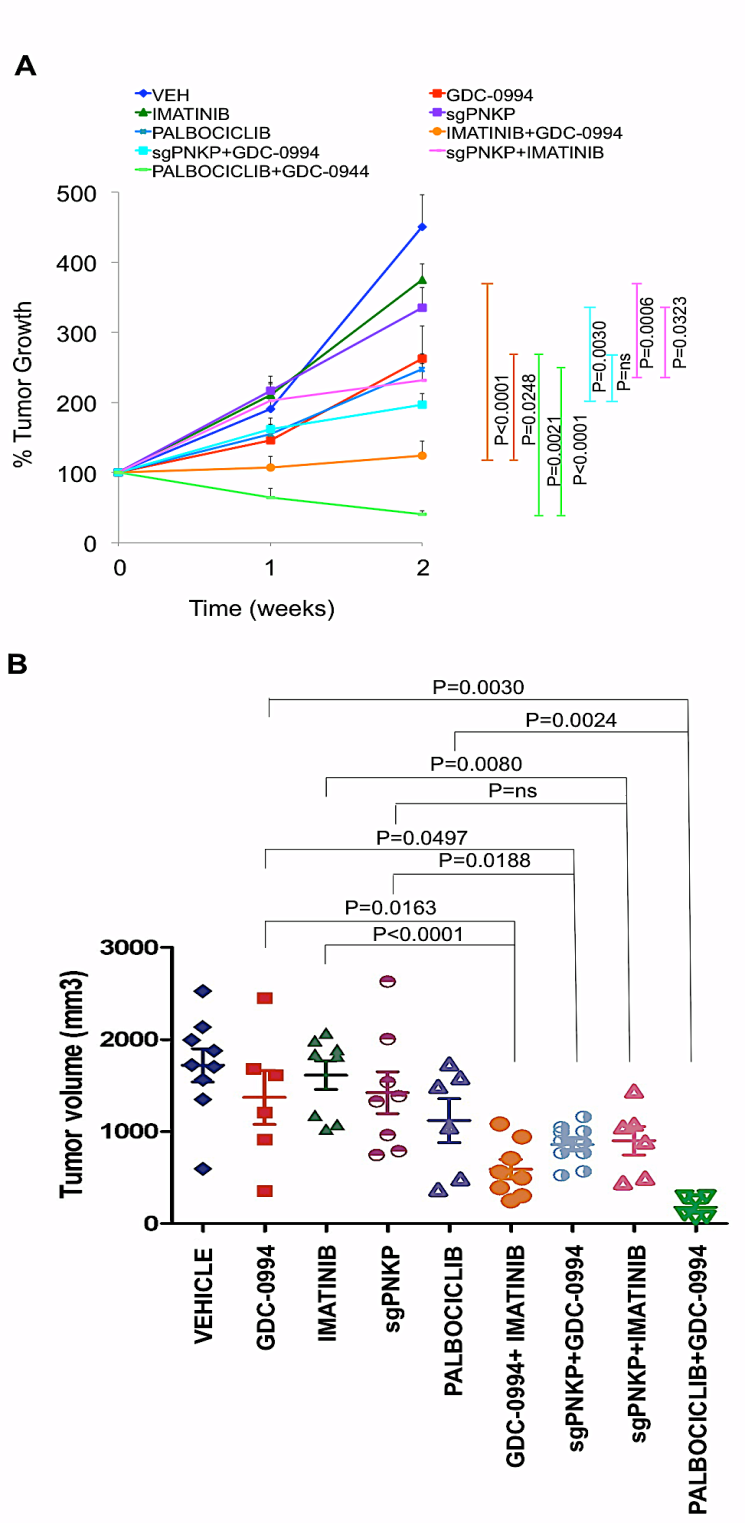


Figure 22. Validation of selected targets in animal models of breast cancer. (A)Tumor growth inhibition was, respectively, 35.5 (GDC-0994), 20.8 (imatinib), 39.6 (palbociclib), 25.6 (sgPNKP), 70.2 (imatinib plus GDC-0994) and 90.0 (palbociclib plus GDC-0994), 55.3 (sgPNKP+imatinib), 56.8 (sgPNKP+GDC-0994) at T=2 weeks. P values are indicated in the figure. (B) Shows tumor burden charts at T=2 weeks. P Values are depicted in the figure. Error bars: SEM.

DISCUSSION

The clinical and molecular heterogeneity of breast cancer is well known. The advancement and widespread application of 'omics' technologies (genomics, epigenomics, transcriptomics or proteomics, among others) has provided unprecedented insights and novel understanding of the molecular complexity of this disease[48,144–147]. In spite of this complexity, clinical decisions still rely primarily on the assessment of three markers: the expression of the endocrine receptors for oestrogen and progesterone (ER and PgR, respectively) and the aberrant expression of HER2. The definition of triple-negative breast cancer (TNBC) applies to all tumors that lack the expression of ER, PgR and HER2, all of which are molecular targets of therapeutic agents[148].

Exploring novel approaches to treatment of these subtypes is critical because less than 30% of women with metastatic breast cancer survive five years and virtually all women with metastatic TNBC will ultimately die of their disease despite systemic therapy. To date, not a single targeted therapy, beyond PARP inhibitors for the BRCA1/2 mutated cases[148], has been approved for the treatment of TNBC and cytotoxic chemotherapy remains the standard treatment.[149]

The introduction of high-throughput technologies that survey thousands of genes and their products in a single assay, coupled with powerful analytical tools, has opened up new avenues for classifying triple-negative breast cancer into biologically and clinically distinct groups based on gene expression patterns [15,17], DNA copy number alterations [147], miRNA expression profiles {Dvinge, 2013 #2} with endnote. Next-generation sequencing (NGS) studies have produced knowledge about this disease and pinpointed novel drivers of TNBC [144,146], however disease segmentation narrowing down the classifiers to the top-29 mutated genes leads to unique set of mutations at the protein level[146]. Despite the efforts of gene-based high-throughput techniques, previous data, unfortunately, have not led to precision-medicine tools yet.

Patient classification tools should not only accurately predict a patient’s clinical course but also provide therapeutic solutions. These results are difficult to achieve because of the nature of high-throughput studies. For instance, the fine-tuning of gene-expression patterns is technically and conceptually challenging. NGS studies have defined cancer

subtypes by point mutations, which are actionable targets in several malignancies, but not in breast cancer[145]. Highly penetrant oncogenes are rarely found in breast cancers, except for HER-2 amplification, which suggests that the "TNBC phenotype" is the result of coexisting, moderately penetrant, genetic changes. A prognostic stratification system directly linked to different therapeutic options is needed for TNBC.

Despite the recognize heterogeneity, TNBC can be simplistically reduced in two clinical subtypes. Tumors of those patients that relapse within 3 years after treatment for logoregional disease, and tumors of other patients, who, with few exceptions rarely relapse[31].

This doctoral thesis aimed to understand further the TNBCs classified by 2 patterns of relapse focusing on discovery of new prognostic and/or predictive markers and novel therapeutic solutions for patients with adverse tumor characteristics. To achieve this aim we chose less commonly used investigation approach named high-throughput phosphoproteomics. This approach was selected firstly because proteins are the “effector molecules of tumor cells[150–154] and secondly, the functional status of proteins is modulated through post-translational modifications (PTMs), of which phosphorylation is the most important[150–154]. The investigation of tumour behaviour by assessing "lower-level -omics" (genomic or transcriptomic profiles) ignores the "upper-level" (proteome) regulatory events[155,156]; thus, interrogating the proteome, and, in particular its PTMs, might provide a more accurate and functional assessment of tumour behaviour. Phosphoproteomic analysis is plagued by the same challenges facing all proteomic experiments: complexity, dynamic range, and temporal dynamics. The true complexity of the phosphoproteome has yet to be determined, but the Phosphosite database (http://www.phosphosite.org) now lists 30 000 phosphorylation sites on 17 000 proteins, and this number is steadily increasing as each large-scale phosphorylation analysis continues to identify a large number of novel sites. With so many of the proteins in the cell being phosphorylated, the dynamic range of the phosphoproteome is similar to that of the proteome (i.e., $1x10^9$), but is further increased by substoichiometric modification. In addition, the temporal dynamics of protein phosphorylation regulate the rapid activation and deactivation of cellular signaling networks, further complicating analysis

of the phosphoproteome. So the challenge is not simply to identify and catalog all of the phosphorylation sites, but rather to identify the site, quantify the stoichiometry, and monitor the temporal change in phosphorylation in response to a variety of cellular perturbations. Performing this task on a large number of phosphorylation sites across a broad swath of the signaling network is especially challenging, but is required to understand the mechanisms by which protein phosphorylation controls cell biology[122].. The application of new proteomic techniques for phosphopeptide quantification is contributing to an improved understanding of cancer cell biology [157–160]. Several techniques for quantitative proteomics have been developed; these can be divided into those that require labeling of proteins with stable isotopes (for example, SILAC and iTRAQ) and those that do not require labeling [161,162]. Approaches based on labeling techniques usually detect a larger number of phosphopeptides than those based on label-free approaches because labeling techniques are compatible with extensive fractionation prior to mass spectrometry analysis. However, because of the time-consuming nature of such analyses, studies based on labeling techniques normally compare a small number of samples with no (or very few) biological replicates, a feature that limit the statistical significance of the results. Therefore there is a trade-off between the number of peptide/proteins identified and samples that can be compared in a study. Label-free approaches are preferred when the aim is to compare large sample numbers and replicates [163]even though the penetrance of the approach may not be as large as when using techniques that allow extensive fractionation before mass spectrometry analysis. Aiming for large sample number analysis we decided that label-free approach is better option for quantification of phosphoproteins in our study. Due to the fact that analysis of phosphoproteom is very demanding[164], prior to initialing such study in tumor biopsies of training set, we first performed few pilot studies involving different sample types. We assessed the performance of Ti^{4+}-IMAC phosphopeptides enrichment strategy in samples from different origin and with different amount of staring material. The number of phosphopeptides in human cell lines and biopsies, and mouse tumor varied depending on the amount of protein material used (Figure 1). We identified ~ 3000 phosphopeptides, in HeLa control sample from 100 µg of protein extract, which is similar as previously shown by Zhou H et all[104] using the same enrichment strategy.

Moreover in the breast cancer cell lines the number of identified phosphopeptides was near 5000 staring from 250µg of protein extract. Increased number of phosphopeptides was expected since the amount of starting material was also higher. The number of phosphopeptides we identified in human biopsies was very similar to the HeLa control, 3000 phosphopeptides, starting from 250µg and only ~500 from 100µg of proteins. Complexity of human tissue complicates the performance of phoshoproteomic strategies, which could explain the lower number of identifications in tumor biopsies compared to cell lines. Since our pilot experiment yielded ferly good results and gave us a global insight regarding number of phosphopeptides that could be identified with selected approach in tumor biopsies and cell lines we proceeded toward high-throughput phosphoproteomic analysis of 34 triple-negative samples from training set and 10 triple-negative cell lines.

1. IDENTIFICATION OF PHOSPHOPROTEOMIC SIGNATURE DRIVING RELAPSING CASES AND AGGRESSIVE CELL LINES.

The mammalian cell constantly receives signals from its surroundings to which it has to respond appropriately. For example, growth-factor signals are integrated with internal-state information and lead to decisions on cell growth, differentiation, or proliferation[165–167]. Many human diseases, including multiple forms of cancer, arise through deregulation of this processing. In last couple of decades, our knowledge of the players in signal transduction mechanisms has been accumulated, mainly through the study of individual molecules in specific pathways. More recently, the emergence of technologies allowing high-throughput experiments such as microarray analysis, has provided a detailed and objective view of downstream transcriptional changes following various stimuli. However, many critical events involved in cellular responses are mediated by changes in posttranslational protein modifications rather than transcriptional changes[168]. A central deficiency in our knowledge of cancer concerns how genomic changes drive the proteome and phosphoproteome to execute phenotypic characteristic[48,147,169,170],

As preciously mentioned triple-negative breast cancer (TNBC), can be simplistically reduced to 2 clinical subtypes: tumours of those patients who relapse within 3 years after treatment for locoregional disease, and tumours of the other patients, who, with few exceptions, rarely relapse[31]. Interrogating the proteome, and, in particular its PTMs, might provide a more accurate and functional assessment of tumour behaviour. We hypothesized that all the existing (genomic, transcriptomic, etc) aberrations across different patients might converge into a discrete number of phosphorylation-driven patterns of activation of the proteome that would account for the 2 main clinical subtypes of TNBC. These patterns would be driven by some of the ~500 kinases encoded by the human genome, and the kinases might be targetable by any more then 200 available clinical-grade inhibitors. Our aims were: to 1) define the phosphoprofiles that could differentiate the relapsed cases from the remainder, 2) identify the hyperactive kinases driving these profiles, 3) translate this information into an application that could be usable in a routine clinical environment, and 4) investigate the usefulness of driver kinases as therapeutic targets. In order to achieve our specific aims we collected a training set of 34 TNBC patients divided in two groups (A and B) based on biphasic pattern of relapse. Of these 34, 13 (group A) were early relapse patients and 21 (group B) late relapse patients (Table1, Figure 2). Moreover in order to later validate our findings in *in vivo*/*in vitro* setup we characterized collection of 10 tnbc cell lines for their aggressive phenotype and preformed phosphoproteomic analysis seeking for matching phospho-signatures. We used astringent model of metastasis, intrperitoneal injection, since the intra-tail vein avoids the process of tumor cells leaving the primary site and entering the bloodstream[171]. As shown in Figure 4 our in vivo study clearly distinguished two groups of cells line - aggressive and indolent, thus mimicking biphasic pattern of relapse behaviour in patients. Three cell lines MDA-MB-231, MDA-MB-157 and Hs578T showed more aggressive behaviour since the animals injected with this cell lines survived less than 100 days. The rest of the cell lines showed more indolent behaviour since the animals survived for more than 400 days. As previously demonstrated by Prat et al., the three cell lines with aggressive behaviour, have a gene expression pattern similar to the claudin-low tumor subtype compared to the other seven cell lines which had gene-expression patter similar to basal-like tumors. In terms

of prognosis, claudin-low tumors showed worse prognosis compared to other subtypes.[17] Moreover recently the study published by Lehmann et all where the authors analyzed gene expression profiles from more than 500 TNBC cases, reported 6 different TNBC subtypes. Two of the subtypes, a mesenchymal (M) and mesenchymal-stem-like (MSL), were enriched in gene expression for epithelial-mesenchymal transition and growth factor pathways and correlated with gene expression of three aggressive cell lines. In the same study the authors showed that the patients with the M and MSL subtype had decreased 5-year survival, consistent with enrichment in pathways associated with metastasis and motility[172]. Thus, the aggressive behavior of MDA-MB-231, MDA-MB-157 AND Hs578T observed in our in vivo experiment is in agreement with previous studies and could be explained by their different genetic features.

Next we did consensus clustering of phosphoproteomic tumor and cell lines profiles and observed a separation between the nature of the phosphopeptides identified in tumor biopsies and cell lines (Figure 5). This difference in phosphoproteomic profile is not that surprising, since we preformed the analysis on established cell lines, which to some extant mirror all the heterogeneity associated with clinical samples. Moreover, complex inter-relationships that exist between cells within the tumor are lost when cell lines are cultured on plastic in two dimensions. Despite this, two-dimensional culture system, still remains the most favored mechanism for *in vitro* studies in breast cancer research. In addition, cell lines are often sensitive to culture conditions - particularly the inclusion of growth factors that can sometimes alter the cell phenotype, resulting in inappropriate pathway activation or differentiation[173,174]. The Venny diagram shown in Figure 7 highlights these differences, with ~1/3 of the phosphosites commonly quantified in tumor biopsies and cell lines. However, gene ontology (GO) analysis performed for the proteins to which identified peptides were mapped, showed enrichment in common pathways with very few GO pathways unique to either the cell lines or the tumor samples Figure 7. Some of the common pathways identified were tight junction, focal adhesion and insulin signaling pathway, that have been already described during the process of adaptation of tumor cell to tissue culture conditions[175]. Although indicating that even from the non-redundant phosphopeptides it

is possible to find the enrichment in common pathways in tumor biopsies and cell lines, this analysis does not consider the functional regulation exerted by the process of phosphorylation. As a matter of fact protein modification sites can serve as nano-switches in the distributed biochemical network of the cell. These switches control the flow of information through a particular protein, or node, of the network, helping to shape dynamic biological processes including genetic silencing, cellular growth, differentiation, and apoptosis[176]. Thus to integrate phosphopeptide status into functionally relevant readouts we performed differential analysis within tumor biopsies and cell lines. The heat-map in Figure 8A shows the 706 peptides identified in tumor biopsies that manifested significantly differential regulation, out of which 161 were hyperphosphorylated in the relapsed cases than in non-relapsed ones. Interestingly, when we searched in the literature for these phosphosites, 63 of them still have not been associated with breast cancer. This highlights the need to conduct the research at this layer of regulation not only for knowledge generation but also because it has been shown that phosphorylation of different sites could be associated with activation of specific nodes[177,178]. Differential analysis in cell lines illustrated with heat-map in Figure 8B showed that 874 peptides, presented significantly differential regulation. Moreover, only 53 peptides were found hyperphosphoryated in the aggressive cell lines compared to indolent ones. Despite the lack of information on the functional description and annotation of many of the identified phosphosites, we sought to interpret the information on phosphosites that we had obtained at a deeper level. From the identified phosphoprofiles we predicted kinases, based on the nature and abundance of phosphorylation events and theoretical affinity of the consensus sequence for each kinase. Furthermore, we preformed gene set enrichment analysis[121] of kinase motifs and we generated enrichment plots for the kinases driving the profiles of relapsed versus non-relapsed cases. Figure 9A shows a chart for kinase PP70S6K, for which increased activity was predicted for the relapsed cases based on the abundance of its substrates. Kinase set enrichment analysis (KSEA) also pinpointed to additional 10 kinase motifs enriched in the aggressive cases. Moreover, the same study preformed in cell lines yielded 3 kinase motifs, PAK2, PRKCE and PIM1, enriched in aggressive cell lines, which were also found in relapsed cases (Figure 9B, C). This result could indicate

some degree of overlap between tumors and cell lines beyond gene ontology (G0) pathways. With the exception of AKT, CDK6, CDK1 and PRKCE, the other kinases have not been previously proposed as drivers of adverse clinical behavior or potential therapeutic targets in TNBC yet. For example overexpressed pAKT was found frequently observed in human lung, gastric, hepatocellular, pancreatic, renal, prostate and endometrial cancer as well as multiple myeloma[179–182]. Studies have documented the prognostic role of pAKT overexpression in some cancers. A recent meta-analysis showed that pAKT overexpression was significantly associated with worse overall survival in non-small cell lung cancer patients (hazard ratio [HR]: 1.38, 95% confidence interval [CI]: 1.11–1.70)[183]. Thus, our finding about the activation of AKT kinase is consistent with the observations in other solid tumors. In the study of Nakanishi et al where authors looked at 135 hepatocellular carcinoma patients, multivariate analysis identified pAKT overexpression as a strong predictor for early disease recurrence (relative risk: 12.5, 95% CI: 2.59–60.55) and poor prognosis (relative risk: 7.90, 95% CI: 1.25–50.00)[180]. The study of Cinti et al in 50 advanced gastric carcinomas showed that the five-year survival rate was 18% in the patients with pAKT overexpression versus 58% in the pAKT-negative ones[184]. These findings together with ours suggest that pAKT overexpression could be a common driver shared by multiple types of human cancer, and thus it has the potential for being a therapeutic target of great clinical significance. Moreover, study done by Hsu et al support our findings since they also linked CDK4/6 as a driver of TNBC. In their study they found eight TNBC-related kinases to be overexpressed in TNBC cell with stem-like properties. Among them, expression of PKC-a, MET, and CDK6 in tumor biopsies correlated with poorer survival outcomes. In cases coexpressing two of these three kinases, survival rates were lower than in cases where only one of these kinases was expressed. In functional tests two-drug combinations targeting these three kinases inhibited TNBC cell proliferation and tumorigenic potential in a cooperative manner. A combination of PKC-a-MET inhibitors also attenuated tumor growth in a cooperative manner in vivo[185]. Another study done by Xia et al analyzed CDK1 as a potential target in breast cancer. The authors investigated CDK1-mediated phosphorylation of BRCA1 and suggested that a reduction of the CDK1 activity by small molecule inhibitors improved the response to PARP inhibition in vitro, and served as a

guide to translate this to substantial antitumor activity in vivo[131]. PRKCE was also previously linked as a driver of TNBC. Data from the cancer genome atlas suggested increased copy number of PRKCE in this adverse subtype of breast cancer. Dann et all showed that overexpression of PKCε in a non-tumorigenic breast epithelial cell line is sufficient to overcome contact inhibition and results in the formation of cellular foci. Correspondingly, inhibition of PKCε in a TNBC cell model results in growth defects in two-dimensional (2D) and three-dimensional (3D) culture conditions and orthotopic xenografts[132]. Thus together with studies previously mentioned, our study highlights that specific kinasis have a very important role in TNBC and are very promising targets for therapeutic intervention.

2. TRANSALTION INTO CLINICAL ASSESSMENTS AND EXTERNAL VALIDATION

In order to rule out that our results were not caused by an imperfect *in silico* algorithm, we sought external validation by analyzing an independent patient series and preclinical *in vitro* and *in vivo* models.

The translation of mass spectrometry into routine clinical practice is currently not straightforward, because of a variety of unresolved issues, including the unsuitability of formalin-fixed paraffin-embedded samples (FFPE) for the technique[133], the effect of time until sample storage on the maintenance of PTMs[134], the logistics of this type of assessment, and cost. So far, two inherent obstacles are usually associated with FFPE tissue proteome initiatives; limited protein extraction due to poor solubility and ambiguous protein identification due to possible unknown peptide modifications. Since formaldehyde is a highly polar compound, a set of complex protein–protein and protein–nucleic acid cross-links are likely to be formed[186,187]. While this covalent complexity preserves the cellular morphology, it constrains sample processing logistics by the less amenable solubility and resistance to a variety of extraction buffers[188,189].Thus, we aimed to translate the information that we obtained by mass spectrometry into data that could be obtained from FFPE samples, while seeking external validation for our results. "Mass spectrometry-to-immunohistochemistry translation" involves 2 main challenges: 1) lack of suitable reagents, meaning that antibodies suitable for use in the

immunohistochemical detection of most of the kinases or upregulated phosphosites identified by phosphoproteomics are not available; 2) unknown criteria for defining an "activated kinase" and identifying the activated form in FFPE samples. Using the algorithm depicted in Figure 10 together with *in vitro* confirmation of the utility of each antibody (Figure 11) we selected the antibodies that we employed for external validation of the hits depicted in Figure 9. This workflow yielded 11 valid antibodies against the following targets: CDK6, CLK1, pS6K (Ser^{240}), DAPK3, PRKCE, pAKT (Ser^{473}), pERK (Thr^{202}/Tyr^{204}), pP70S6K (Thr^{389}), PIM1, pPNKP (Ser^{114}/Thr^{118}) and c-KIT. The 11 targets were stained in an independent set of TNBCs of 113 patients which clinical and demographic characteristics and Kaplan-Meier curve are presented in Table 2 and Figure 12, respectively. Advantages of automated IHC measurements over pathologist visual scoring are precision in ranges of staining that appear weak to the eye[190] and production of continuous data. Moreover, when automated IHC measurements are provided to a pathologist during visual scoring, computer-aided IHC analysis substantially improves both intra- and inter-observer agreement[191]. For these reasons we selected automated quantification of the positive staining and generation of IHC-scores for each antibody. IHC-scores we than divided into quartiles (Table 3, Figure13) for subsequent prognostic analysis. We considered a kinase of interest “highly active” if IHC-score was in the upper quartile. When using upper quartile staining, for each kinase, to stratify patients by their time to relapse we found that 6 out of 11 kinases were able to predict the clinical course of patients in the validation set, as shown by Kaplan-Meier curves and the hazard rations (Figure 14), which were adjusted by T, N, grade (G) and age. These results appear to be mostly novel since reports linking any of the 6 kinases to prognosis or therapeutic potential in TNBC are very scars. ERK has been reported to be associated with bad prognosis and suggested as a potential target. For example Bartholomeusz et all showed that high ERK (ERK-2 and pMAPK) expression is an important predictor of OS and disease-free survival outcomes in patients with TNBC and suggested that ERK expression may be the driving force in TNBC[192]. This is also in agreement with study done by Fujii et all where the authors concluded that ERK has a pleiotropic role in gene regulation by controlling the phosphorylation of transcription factors, including Elk-1, and chromatin remodeling via

EZH2 and histone H3 methylation. These effects seemed to be critical to the aggressiveness and high growth rate of triple-negative breast cancer[193].

In the case of other kinases such as CDK6, as mentioned previously Hsu et all reported that high CDK6 function as a potential factor for adverse clinical course[185], while the reports regarding its potential role as a target in TNBC are contradictory[185,194]. Similar results we found for the c-KIT, where Burtain et all reported the amplification of c-KIT in a subgroup of TNBC patients with a intermediate prognosis[195], (while no positive results as a potential target exist). However, PNKP, PRKCE and P70S6K protein have not been linked yet to either prognosis or targetability in this disease. Finally lack of significant co-linearity between kinases (Figure 15) and discrete number of activated patterns (Figure 16) per patient leaded us to investigate prognostic information of 6 kinases together. When one or more of them concurrently co-exist, the relapse risk increases greater than 9-fold risk of relapse, even when adjusted for the conventional risk factors (Figure 17). Eventhough each of the 6 kinases confers on itself clinically useful information (Figure 14), the lack of relapse in a group of patients without any of the 6 kinases active, allowed us to detect which of the signalling nodes seam to be sufficient to maintain the biology of the aggressive cases. According to the status (high activity or not) of the kinases in the tumours from the validation set, we observed two main patterns. One, without high activity among any of the 6 kinases that was associated with a very high long-term relapse-free rate. From the clinical point of view the ability to identify patients belonging to this group is very important, since therapeutic efforts could be focused on other patients not in this group. The second pattern (with 37 sub-patterns) was characterized by the hyperactivity of one or more kinases; all but two of the patients who developed relapse had tumours showing this pattern (Figure 16, Table 4). These results suggest that the activation of any of these kinases was sufficient to cause an aggressive clinical course. To the best of our knowledge, we are not aware of any other prognostic tool that can stratify the outcomes of patients with TNBC with the accuracy of K-high. Interestingly, the 6 kinases of K-high can be measured rapidly and inexpensively. More importantly, in the context of the lack of drivers of oncogene addiction in TNBC, our strategy narrows down the potential millions of two-by-two drug screening combinations of currently existing targeted drugs, to a 15 possible

combinations with agents that target those 6 targets for TNBC patients that are not cured by the current standard of care. The 15 drug combinations, which were tested in 10 triple-negative cell lines, showed supra-additive effect in virtually 100% of the tested cases (Figure 19A A,B). The drug combinations that showed highest synergistic effect were tested also in animal models of breast cancer. As shown in Figure 22 tumor growth inhibition was significantly higher for any of the tested combinations compared to single agents that demonstrated models effect both *in vitro* (Figure 17 and 18) and *in vivo* Figure 22. Although future research should be conducted to confirming that the de-regulated signalling pathways in tumours are similar to those in tumour cell lines, we consider that our therapeutic results are valid, since we found that the GO pathways (Figure 7) and driver kinases (Figure 9) overlapped between the tumour cell lines and patient tumour samples. Altogether, our results suggest that many different combinations of elevated output in such disparate pathways as "translational control" (pP70S6K), "calcium signaling" (PRKCE), "cell replication" (CDK6), "growth factor signaling" (c-KIT), "DNA damage" (pPNKP) and "MAP-kinases" (pERK1/2) lead to an aggressive, although potentially therapeutically targetable course of disease in TNBC.

CONCLUSIONS

We have accumulated compelling evidence indicating that phosphoproteomic profiling of triple-negative breast cancer could provide both identification of subgroups with different clinical outcome and more importantly novel therapeutic solution for patients with adverse clinical prognosis. In summary we found that:

1. High-through mass-spectrometry based phosphoproteomics is a powerful tool that allows gain of new knowledge by identification of phosphorylated proteins that are not previously linked to triple-negative breast cancer.
2. Phosphorylated proteins identified in triple-negative breast cancer cases and cell lines are intrinsically different and cluster into two different clusters, however they could be reduced to a much smaller number of overlapping driver kinases.
3. Kinase set enrichment analysis pinpointed significantly enriched kinases driving the profiles of aggressive triple-negative breast cancer cases.
4. Translation of mass-spectrometry into immunohistochemistry, even thought technically demanding, is feasible and provides generation of prognostic and/or predictive signatures.
5. We identified six kinases, PRKCE, P70S6K, PNKP, ERK, c-KIT and CDK6 which increased activity appears to drive aggressive clinical course of TNBC.
6. Study of the status of each of six kinases in the tumors from the validation set allowed us the assembly of a basic kinome landscape of TNBC. One of the found patterns was characterized by the lack of high activity in the six kinases and was associated with a very high long-term relapse-free rate. Other pattern in which one or more kinases showed hyperactivation was associated with aggressive clinical course.
7. Hyperactivity of one or more kinases confers greater than 9-fold risk of relapse, even when adjusted for the conventional risk factors. Moreover K-high is able to stratify the outcome of patients with TNBC better than any currently available prognostic tool.

8. Our strategy allowed the design of potential two-by-two drug screening combinations with currently existing targeted drugs. The 15 doublets showed supra-additive effect in almost all triple-negative cell lines tested.
9. Targeting ERK, c-KIT, CDK6 and PNKP in MDA-MB-231 aggressive cell lines showed modest therapeutic potential. However drug combinations demonstrate significant reduction of tumor growth.
10. The identified drug combinations may serve as novel therapeutic solutions for patients positive for K-high although further validation in clinical trials should be conducted.

CONCLUSIONES

Hemos llegado a evidencias convincentes que nos indican que los perfiles de fosforilacion de cáncer de mama triple negativo podrían proporcionarnos una identificación de los distintos subgrupos en base a su evolución clínica y más importante proponer una solución terapéutica para los pacientes con pronóstico clínico adverso. En resumen se encontró que:

1. El estudio de la fosfoproteómica basado en técnicas de espectrometría de masas es una herramienta útil que nos permite, mediante la identificacion de nuevas proteínas fosforiladas que previamente no estaban vinculadas al cáncer de mama triple negarivo aumentar el conocimiento que tenemos de este tipo de cáncer
2. Las proteínas fosforiladas identificadas en las muestras de pacientes de cáncer de mama triple negativo y las identificadas en las líneas celulares son diferentes y se agrupan en dos grupos diferentes, sin embargo estas podrían reducirse a un número mucho menor si solo tenemos en cuenta las que están presentes a la vez en ambos grupos.
3. Mediante el estudio de enriquecimiento de fosfopéptidos se identificaron kinasas que estaban significativamente enriquecidas en los casos de cáncer de cancer de mama con recaida y que podrían conducir a este fenotipo agresivo.
4. Aplicar los datos obtenidos mediante espectometría de masas mediante técnicas inmunohistológicas es factible y proporciona la generación de nuevos marcadores de pronóstico y predictivos.
5. Se han identificado seis kinasas: PRKCE, P70S6K, PNKP, ERK, c-KIT y CDK6 cuya mayor actividad es capaz de dirigir el curso clínico agresivo del TNBC.
6. El estudio del estatus de cada una de las seis kinasas identificadas usando el set de validación nos permitió tener una visión general del quinoma básico presente en los casos de TNBC. Uno de los patrones encontrados se caracterizó por la falta de actividad alta en las seis kinasas y se asoció con una probabilidad muy alta de no recaida a largo plazo. Sin embargo, el otro patrón en el que una o más de las kinasas identificadas presentaban valores de hiperactivatición se asoció con un curso clínico agresivo.

7. La hiperactividad de una o varias de las kinasas identificadas confiere un mayor riesgo de recaída (9 veces mayor). Por otra parte alta-K es capaz de estratificar el comportamiento de los pacientes con TNBC mejor que cualquier herramienta de pronóstico actualmente disponible.
8. Nuestra estrategia permitío el diseño de 2 por 2 combinaciones con farmacos existentes actualmente. Los 15 dobletes mostraron un efecto aditivo en casi todas las celulas líneares triple negativas probadas.
9. El uso de fármacos dirigidos contra ERK, c-KIT, CDK6 y PNKP en las líneas celulares agresivas, MDA-MB-231, demostró un modesto potencial terapéutico. Sin embargo, el uso de las combinaciones de fármacos demostraron una reducción significativa del crecimiento tumoral.
10. Las combinaciones de medicamentos identificadas pueden servir como nuevas soluciones terapéuticas para pacientes positivos de alta-K aunque es necesarió llevar a cabo ensayos clínicos que validen los resultados obtenidos.

REFERENCES

1. Kakarala, M. & Wicha, M. S. Implications of the cancer stem-cell hypothesis for breast cancer prevention and therapy. *J Clin Oncol* **26,** 2813–2820 (2008).

2. Taborga, M., Corcoran, K. E., Fernandes, N., Ramkissoon, S. H. & Rameshwar, P. G-coupled protein receptors and breast cancer progression: potential drug targets. *Mini Rev. Med. Chem.* **7,** 245–251 (2007).

3. Maxmen, A. Statistics: The hard facts. *Nature* **485,** S50–S51 (2012).

4. Lacey Jr., J. V *et al.* Breast cancer epidemiology according to recognized breast cancer risk factors in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial Cohort. *BMC Cancer* **9,** 84 (2009).

5. Hankinson, S. E., Colditz, G. a & Willett, W. C. Towards an integrated model for breast cancer etiology: the lifelong interplay of genes, lifestyle, and hormones. *Breast Cancer Res.* **6,** 213–8 (2004).

6. Fackenthal, J. D. & Olopade, O. I. Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. *Nat. Rev. Cancer* **7,** 937–948 (2007).

7. Gasco, M., Shami, S. & Crook, T. The p53 pathway in breast cancer. *Breast Cancer Res.* **4,** 70–76 (2002).

8. Guinee, V. F. *et al.* Effect of pregnancy on prognosis for young women with breast cancer. *Lancet (London, England)* **343,** 1587–9 (1994).

9. DeBruin, L. S. & Josephy, P. D. Perspectives on the chemical etiology of breast cancer. *Environ. Health Perspect.* **110,** 119–128 (2002).

10. Malhotra, G. K., Zhao, X., Band, H. & Band, V. Histological, molecular and functional subtypes of breast cancers. *Cancer Biology and Therapy* **10,** 955–960 (2010).

11. Kumar, V., Fausto, N., Abbas, A. K. & Aster, J. C. *Robbins & Cotran Pathologic Basis of Disease*. *Journal of Clinical Pathology* (2010). doi:citeulike-article-id:4745943

12. Eheman, C. R. *et al.* The changing incidence of in situ and invasive ductal and lobular breast carcinomas: United States, 1999-2004. *Cancer Epidemiol. Biomarkers Prev.* **18,** 1763–1769 (2009).

13. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406,** 747–752 (2000).

14. Rosai J. *Rosai and Ackerman's Surgical Pathology*. *ed. MOSBY Elsewier* (2011). doi:10.1016/B978-0-323-06969-4.00001-5

15. Herschkowitz, J. I. *et al.* Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol.* **8,** R76 (2007).

16. Perou, C. M. Molecular stratification of Triple negative breast cancer. *The on* **15,** 39–48 (2010).

17. Prat, A. *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast cancer Res.* **12,** R68 (2010).

18. Creighton, C. J. *et al.* Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 13820–13825 (2009).

19. Haffty, B. G. *et al.* Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer. *J. Clin. Oncol.* **24,** 5652–5657 (2006).

20. Rakha, E. A. *et al.* Prognostic markers in triple-negative breast cancer. *Cancer* **109,** 25–32 (2007).

21. Dent, R. *et al.* Triple-negative breast cancer: Clinical features and patterns of recurrence. *Clin. Cancer Res.* **13,** 4429–4434 (2007).

22. Anders, C. K. & Carey, L. A. Biology, metastatic patterns, and treatment of patients with triple-negative breast cancer. *Clin. Breast Cancer* **9 Suppl 2,** S73-81

(2009).

23. Reis-Filho, J. S. & Tutt, A. N. J. Triple negative tumours: A critical review. *Histopathology* **52,** 108–118 (2008).

24. Nanda, R. 'Targeting' triple-negative breast cancer: The lessons learned from BRCA1-associated breast cancers. *Semin. Oncol.* **38,** 254–262 (2011).

25. Phipps, A. I. *et al.* Body size, physical activity, and risk of triple-negative and estrogen receptor-positive breast cancer. *Cancer Epidemiol. Biomarkers Prev.* **20,** 454–63 (2011).

26. Thike, A. A. *et al.* Triple-negative breast cancer: clinicopathological characteristics and relationship with basal-like breast cancer. *Mod. Pathol.* **23,** 123–133 (2009).

27. Thike, A. A. *et al.* Triple negative breast cancer: outcome correlation with immunohistochemical detection of basal markers. *Am. J. Surg. Pathol.* **34,** 956–64 (2010).

28. Morris, G. J. *et al.* Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients: A single-institution compilation compared with the national cancer institute's surveillance, epidemiology, and end results database. *Cancer* **110,** 876–884 (2007).

29. Carey, L. A. *et al.* The triple negative paradox: Primary tumor chemosensitivity of breast cancer subtypes. *Clin. Cancer Res.* **13,** 2329–2334 (2007).

30. Tischkowitz, M. *et al.* Use of immunohistochemical markers can refine prognosis in triple negative breast cancer. *BMC Cancer* **7,** 134 (2007).

31. Foulkes, W. D., Smith, I. E. & Reis-Filho, J. S. Triple-Negative Breast Cancer. *N. Engl. J. Med.* **363,** 1938–1948 (2010).

32. Banerjee, S. *et al.* Basal-like breast carcinomas: clinical outcome and response to chemotherapy. *J. Clin. Pathol.* **59,** 729–35 (2006).

33. Rakha, E. A. *et al.* Are triple-negative tumours and basal-like breast cancer

synonymous? *Breast Cancer Res.* **9,** 404; author reply 405 (2007).

34. Rakha, E. A., El-Sayed, M. E., Reis-Filho, J. & Ellis, I. O. Patho-biological aspects of basal-like breast cancer. *Breast Cancer Research and Treatment* **113,** 411–422 (2009).

35. Nielsen, T. O. *et al.* Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin. Cancer Res.* **10,** 5367–5374 (2004).

36. Bertucci, F. *et al.* How basal are triple-negative breast cancers? *Int. J. Cancer* **123,** 236–240 (2008).

37. Weigelt, B., Baehner, F. L. & Reis-Filho, J. S. The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: A retrospective of the last decade. *Journal of Pathology* **220,** 263–280 (2010).

38. Rakha, E. A., Reis-Filho, J. S. & Ellis, I. O. Impact of basal-like breast carcinoma determination for a more specific therapy. *Pathobiology* **75,** 95–103 (2008).

39. Calza, S. *et al.* Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. *Breast Cancer Res.* **8,** R34 (2006).

40. De Ronde, J. J. *et al.* Concordance of clinical and molecular breast cancer subtyping in the context of preoperative chemotherapy response. *Breast Cancer Res. Treat.* **119,** 119–126 (2010).

41. Bernard, P. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27,** 1160–1167 (2009).

42. Ashman, K. & Villar, E. L. Phosphoproteomics and cancer research. *Clin. Transl. Oncol.* **11,** 356–362 (2009).

43. Weigelt, B. *et al.* Breast cancer molecular profiling with single sample predictors: A retrospective analysis. *Lancet Oncol.* **11,** 339–349 (2010).

44. Hu, Z. Insight into microRNA regulation by analyzing the characteristics of their

targets in humans. *BMC Genomics* **10,** 594 (2009).

45. Hennessy, B. T. *et al.* Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. *Cancer Res.* **69,** 4116–4124 (2009).

46. Rakha, E. A. *et al.* Triple-negative breast cancer: Distinguishing between basal and nonbasal subtypes. *Clin. Cancer Res.* **15,** 2302–2310 (2009).

47. Economopoulou, P., Dimitriadis, G. & Psyrri, A. Beyond BRCA: New hereditary breast cancer susceptibility genes. *Cancer Treatment Reviews* **41,** 1–8 (2015).

48. Network, T. C. G. A. Comprehensive molecular portraits of human breast tumors. *Nature* **490,** 61–70 (2012).

49. Stephens, P. J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486,** 400–4 (2012).

50. Langerod, A. AL *et al.* TP53 mutation status and gene expression profiles are powerful prognostic markers of breast cancer. *Breast Cancer Res.* **9,** R30 (2007).

51. Børresen-Dale, A. L. TP53 and breast cancer. *Human Mutation* **21,** 292–300 (2003).

52. Oakman, C., Viale, G. & Di Leo, A. Management of triple negative breast cancer. *Breast* **19,** 312–21 (2010).

53. Sorlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 8418–23 (2003).

54. Carey, L., Winer, E., Viale, G., Cameron, D. & Gianni, L. Triple-negative breast cancer: disease entity or title of convenience? *Nat. Rev. Clin. Oncol.* **7,** 683–692 (2010).

55. Kennedy, R. D., Quinn, J. E., Mullan, P. B., Johnston, P. G. & Harkin, D. P. The role of BRCA1 in the cellular response to chemotherapy. *Journal of the National Cancer Institute* **96,** 1659–1668 (2004).

56. Turner, N., Tutt, A. & Ashworth, A. Targeting the DNA repair defect of BRCA tumours. *Current Opinion in Pharmacology* **5,** 388–393 (2005).

57. Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434,** 917–21 (2005).

58. Liedtke, C. *et al.* Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J. Clin. Oncol.* **26,** 1275–81 (2008).

59. Chacón, R. D. & Costanzo, M. V. Triple-negative breast cancer. *Breast Cancer Res.* **12 Suppl 2,** S3 (2010).

60. von Minckwitz, G. & Martin, M. Neoadjuvant treatments for triple-negative breast cancer (TNBC). *Ann. Oncol.* **23,** (2012).

61. Hugh, J. *et al.* Breast cancer subtypes and response to docetaxel in node-positive breast cancer: Use of an immunohistochemical definition in the BCIRG 001 trial. *J. Clin. Oncol.* **27,** 1168–1176 (2009).

62. Thomas, E. S. *et al.* Ixabepilone plus capecitabine for metastatic breast cancer progressing after anthracycline and taxane treatment. *J. Clin. Oncol.* **25,** 5210–5217 (2007).

63. Cleator, S., Heller, W. & Coombes, R. C. Triple-negative breast cancer: therapeutic options. *Lancet Oncol.* **8,** 235–244 (2007).

64. Troester, M. A. *et al.* Gene expression patterns associated with p53 status in breast cancer. *BMC Cancer* **6,** 276 (2006).

65. Moyano, J. V. *et al.* αB-Crystallin is a novel oncoprotein that predicts poor clinical outcome in breast cancer. *J. Clin. Invest.* **116,** 261–270 (2006).

66. Sparano, J. A. *et al.* GRB7-dependent pathways are potential therapeutic targets in triple-negative breast cancer. *Cancer Res.* **69,** (2009).

67. A, A. A. M. J. H. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **14,** 35–48 (2012).

68. Sallam, R. M. Proteomics in cancer biomarkers discovery: Challenges and applications. *Disease Markers* **2015,** (2015).

69. Wang, M., You, J., Bemis, K. G., Tegeler, T. J. & Brown, D. P. G. Label-free mass spectrometry-based protein quantification technologies in proteomic analysis. *Briefings in Functional Genomics and Proteomics* **7,** 329–339 (2008).

70. Balabanov, S. *et al.* Tumour-related enzyme alterations in the clear cell type of human renal cell carcinoma identified by two-dimensional gel electrophoresis. *Eur. J. Biochem.* **268,** 5977–5980 (2001).

71. Shi, T. *et al.* Differential protein profiling in renal-cell carcinoma. *Mol. Carcinog.* **40,** 47–61 (2004).

72. Zhu, W., Smith, J. W. & Huang, C.-M. Mass spectrometry-based label-free quantitative proteomics. *J. Biomed. Biotechnol.* **2010,** 840518 (2010).

73. Chen, E. I. & Yates, J. R. Cancer proteomics by quantitative shotgun proteomics. *Molecular Oncology* **1,** 144–159 (2007).

74. Veenstra, T. D. Global and targeted quantitative proteomics for biomarker discovery. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* **847,** 3–11 (2007).

75. Voyksner, R. D. & Lee, H. Investigating the use of an octupole ion guide for ion storage and high-pass mass filtering to improve the quantitative performance of electrospray ion trap mass spectrometry. *Rapid Commun. Mass Spectrom.* **13,** 1427–1437 (1999).

76. Chelius, D. & Bondarenko, P. V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J. Proteome Res.* **1,** 317–323 (2002).

77. Bondarenko, P. V., Chelius, D. & Shaler, T. A. Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography - Tandem mass spectrometry. *Anal.*

Chem. **74,** 4741–4749 (2002).

78. Lengqvist, J. *et al.* Robustness and accuracy of high speed LC-MS separations for global peptide quantitation and biomarker discovery. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **877,** 1306–1316 (2009).

79. Fatima, N. *et al.* Label-free global serum proteomic profiling reveals novel celecoxib-modulated proteins in familial adenomatous polyposis patients. *Cancer Genomics and Proteomics* **6,** 41–50 (2009).

80. Huang, S. K. *et al.* LC/MS-based quantitative proteomic analysis of paraffin-embedded archival melanomas reveals potential proteomic biomarkers associated with metastasis. *PLoS One* **4,** (2009).

81. Levin, Y., Schwarz, E., Wang, L., Leweke, F. M. & Bahn, S. Label-free LC-MS/MS quantitative proteomics for large-scale biomarker discovery in complex samples. *J. Sep. Sci.* **30,** 2198–2203 (2007).

82. Vissers, J. P. C., Langridge, J. I. & Aerts, J. M. F. G. Analysis and quantification of diagnostic serum markers and protein signatures for Gaucher disease. *Mol. Cell. proteomics* **6,** 755–66 (2007).

83. Huang, J. T. J. *et al.* CSF biomarker discovery using label-free nano-LC-MS based proteomic profiling: Technical aspects. *J. Sep. Sci.* **30,** 214–225 (2007).

84. Washburn, M. P., Wolters, D. & Yates 3rd, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19,** 242–247 (2001).

85. Liu, H., Sadygov, R. G. & Yates 3rd, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76,** 4193–4201 (2004).

86. Pang, J. X., Ginanni, N., Dongre, A. R., Hefta, S. A. & Opiteck, G. J. Biomarker discovery in urine by proteomics. *J. Proteome Res.* **1,** 161–169 (2002).

87. Rao, P. V. *et al.* Proteomic identification of salivary biomarkers of type-2 diabetes. *J. Proteome Res.* **8,** 239–245 (2009).

88. Old, W. M. *et al.* Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* **4,** 1487–1502 (2005).

89. Zybailov, B., Coleman, M. K., Florens, L. & Washburn, M. P. Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Anal. Chem.* **77,** 6218–6224 (2005).

90. Florens, L. *et al.* Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **40,** 303–311 (2006).

91. Pan, J., Chen, H.-Q., Sun, Y.-H., Zhang, J.-H. & Luo, X.-Y. Comparative proteomic analysis of non-small-cell lung cancer and normal controls using serum label-free quantitative shotgun technology. *Lung* **186,** 255–61 (2008).

92. Asara, J. M., Christofk, H. R., Freimark, L. M. & Cantley, L. C. A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics* **8,** 994–999 (2008).

93. Seyfried, N. T. *et al.* Up-regulation of NG2 proteoglycan and interferon-induced transmembrane proteins 1 and 3 in mouse astrocytoma: A membrane proteomics approach. *Cancer Lett.* **263,** 243–252 (2008).

94. Huber, K. V. M. & Superti-Furga, G. in *Methods in Molecular Biology* **1394,** 211–218 (2016).

95. Zhang, J., Yang, P. L. & Gray, N. S. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **9,** 28–39 (2009).

96. Choudhary, C. & Mann, M. Decoding signalling networks by mass spectrometry-based proteomics. *Nat. Rev. Mol. Cell Biol.* **11,** 427–439 (2010).

97. Hein, M. Y., Sharma, K., Cox, J. & Mann, M. *Chapter 1 - Proteomic Analysis of Cellular Systems*. *Handbook of Systems Biology* (2013).

98. J??nger, M. A. & Aebersold, R. Mass spectrometry-driven phosphoproteomics: Patterning the systems biology mosaic. *Wiley Interdisciplinary Reviews: Developmental Biology* **3,** 83–112 (2014).

99. Lemeer, S. & Heck, A. J. The phosphoproteomics data explosion. *Current Opinion in Chemical Biology* **13,** 414–420 (2009).

100. Echols, N. *et al.* Comprehensive analysis of amino acid and nucleotide composition in eukaryotic genomes, comparing genes and pseudogenes. *Nucleic Acids Res.* **30,** 2515–23 (2002).

101. Ubersax, J. a & Ferrell, J. E. Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* **8,** 530–41 (2007).

102. Humphrey, S. J. *et al.* Dynamic adipocyte phosphoproteome reveals that akt directly regulates mTORC2. *Cell Metab.* **17,** 1009–1020 (2013).

103. Lundby, A. *et al.* Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues. *Nat. Commun.* **3,** 876 (2012).

104. Zhou, H. *et al.* Toward a comprehensive characterization of a human cancer cell phosphoproteome. *J. Proteome Res.* **12,** 260–271 (2013).

105. Boersema, P. J. *et al.* In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling. *Mol. Cell. Proteomics* **9,** 84–99 (2010).

106. Guo, Z. *et al.* Regulation of androgen receptor activity by tyrosine phosphorylation. *Cancer Cell* **10,** 309–319 (2006).

107. Wu, X. *et al.* Phosphoproteomic analysis identifies FAK2 as a potential therapeutic target for tamoxifen resistance in breast cancer. *Mol. Cell. Proteomics*

mcp.M115.050484 (2015). doi:10.1074/mcp.M115.050484

108. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug. Discov.* **5,** 993–996 (2006).

109. Srinivas, P. R., Kramer, B. S. & Srivastava, S. Trends in biomarker research for cancer detection. *Lancet Oncol.* **2,** 698–704 (2001).

110. Goss, V. L. *et al.* A common phosphotyrosine signature for the Bcr-Abl kinase. *Blood* **107,** 4888–4897 (2006).

111. Zhou, H. *et al.* Enhancing the identification of phosphopeptides from putative basophilic kinase substrates using Ti (IV) based IMAC enrichment. *Mol Cell Proteomics* **10,** M110 006452 (2011).

112. de Graaf, E. L., Giansanti, P., Altelaar, A. F. M. & Heck, A. J. R. Single-step enrichment by Ti4+-IMAC and label-free quantitation enables in-depth monitoring of phosphorylation dynamics with high reproducibility and temporal resolution. *Mol. Cell. Proteomics* **13,** 2426–34 (2014).

113. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6,** 359–62 (2009).

114. Zhou, H. *et al.* Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat. Protoc.* **8,** 461–80 (2013).

115. Frese, C. K. *et al.* Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J. Proteome Res.* **10,** 2377–2388 (2011).

116. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–72 (2008).

117. Amanchy, R. *et al.* A curated compendium of phosphorylation motifs. *Nat.*

Biotechnol. **25,** 285–286 (2007).

118. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14,** 128 (2013).

119. Kuleshov, M. V *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44,** W90-7 (2016).

120. Smyth, G. K., Michaud, J. & Scott, H. S. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21,** 2067–2075 (2005).

121. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15545–50 (2005).

122. Arias-Romero, L. E. *et al.* A Rac-Pak signaling pathway is essential for ErbB2-mediated transformation of human breast epithelial cancer cells. *Oncogene* **29,** 5839–5849 (2010).

123. Myhre, S. *et al.* Influence of DNA copy number and mRNA levels on the expression of breast cancer related proteins. *Mol. Oncol.* **7,** 704–718 (2013).

124. Park, E. S. *et al.* Integrative analysis of proteomic signatures, mutations, and drug responsiveness in the NCI 60 cancer cell line set. *Mol. Cancer Ther.* **9,** 257–67 (2010).

125. Shankavaram, U. T. *et al.* Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study. *Mol. Cancer Ther.* **6,** 820–32 (2007).

126. Bonora, M. *et al.* Molecular mechanisms of cell death: central implication of ATP synthase in mitochondrial permeability transition. *Oncogene* **34,** 1608 (2015).

127. L??pez, E. & Cho, W. C. S. Phosphoproteomics and lung cancer research. *International Journal of Molecular Sciences* **13,** 12287–12314 (2012).

128. Xiao, H. *et al.* Proteomic analysis of human saliva from lung cancer patients using two-dimensional difference gel electrophoresis and mass spectrometry. *Mol Cell Proteomics* **11,** M111 012112 (2012).

129. Balko, J. M. *et al.* Molecular profiling of the residual disease of triple-negative breast cancers after neoadjuvant chemotherapy identifies actionable therapeutic targets. *Cancer Discov.* **4,** 232–245 (2014).

130. Dean, J. L., McClendon, A. K. & Knudsen, E. S. Modification of the DNA damage response by therapeutic CDK4/6 inhibition. *J. Biol. Chem.* **287,** 29075–29087 (2012).

131. Xia, Q. *et al.* The CDK1 inhibitor RO3306 improves the response of BRCA-proficient breast cancer cells to PARP inhibition. *Int. J. Oncol.* **44,** 735–744 (2014).

132. Dann, S. G. *et al.* p120 catenin is a key effector of a Ras-PKCε oncogenic signaling axis. *Oncogene* **33,** 1385–94 (2014).

133. Magdeldin, S. & Yamamoto, T. Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues. *Proteomics* **12,** 1045–1058 (2012).

134. Mertins, P. *et al.* Ischemia in tumors induces early and sustained phosphorylation changes in stress kinase pathways but does not affect global protein levels. *Mol Cell Proteomics* **13,** 1690–1704 (2014).

135. Alessi, D. R. *et al.* Mechanism of activation of protein kinase B by insulin and IGF-1. *EMBO J.* **15,** 6541–51 (1996).

136. Warfel, N. A. & Kraft, A. S. PIM kinase (and Akt) biology and signaling in tumors. *Pharmacol. Ther.* **151,** 41–49 (2015).

137. Johnson, J. E., Giorgione, J. & Newton, A. C. The C1 and C2 domains of protein kinase C are independent membrane targeting modules, with specificity for phosphatidylserine conferred by the C1 domain. *Biochemistry* **39,** 11360–11369 (2000).

138. Cenni, V. *et al.* Regulation of novel protein kinase C epsilon by phosphorylation. *Biochem. J.* **363,** 537–45 (2002).

139. Hirota, S. *et al.* Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science* **279,** 577–580 (1998).

140. Lasota, J., Jasinski, M., Sarlomo-Rikala, M. & Miettinen, M. Mutations in exon 11 of c-Kit occur preferentially in malignant versus benign gastrointestinal stromal tumors and do not occur in leiomyomas or leiomyosarcomas. *Am. J. Pathol.* **154,** 53–60 (1999).

141. Gorin, M. A. & Pan, Q. Protein kinase C epsilon: an oncogene and emerging tumor biomarker. *Mol. Cancer* **8,** 9 (2009).

142. Sattler, M. *et al.* Steel factor induces tyrosine phosphorylation of CRKL and binding of CRKL to a complex containing c-Kit, phosphatidylinositol 3-kinase, and p120(CBL). *J. Biol. Chem.* **272,** 10248–10253 (1997).

143. Blume-Jensen, P. *et al.* Kit/stem cell factor receptor-induced activation of phosphatidylinositol 3'-kinase is essential for male fertility. *Nat. Genet.* **24,** 157–62 (2000).

144. Banerji, S. *et al.* Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* **486,** 405–9 (2012).

145. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **503,** 333–339 (2013).

146. Shah, S. P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486,** 395–9 (2012).

147. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2 , 000 breast tumours. *Nature* 1–7 (2012). doi:10.1038/nature10983

148. Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E. & Gianni, L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease.

Nat. Rev. Clin. Oncol. **13,** 674–690 (2016).

149. Mayer, I. A., Abramson, V. G., Lehmann, B. D. & Pietenpol, J. A. New strategies for triple-negative breast cancer-deciphering the heterogeneity. *Clin. Cancer Res.* **20,** 782–790 (2014).

150. Piggee, C. Phosphoproteomics: Miles to go before it's routine. *Anal. Chem.* **81,** 2418–2420 (2009).

151. Thingholm, T. E., Jensen, O. N. & Larsen, M. R. Analytical strategies for phosphoproteomics. *Proteomics* **9,** 1451–1468 (2009).

152. Grimsrud, P. A., Swaney, D. L., Wenger, C. D., Beauchene, N. A. & Coon, J. J. Phosphoproteomics for the masses. *ACS Chemical Biology* **5,** 105–119 (2010).

153. Stasyk, T. & Huber, L. A. Mapping in vivo signal transduction defects by phosphoproteomics. *Trends in Molecular Medicine* **18,** 43–51 (2012).

154. Dissmeyer, N. & Schnittger, A. The age of protein kinases. *Methods in Molecular Biology* **779,** 7–52 (2011).

155. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 1–20 (2016). doi:10.1038/nature17676

156. Bossi, D. *et al.* In Vivo Genetic Screens of Patient-Derived Tumors Revealed Unexpected Frailty of the Transformed Phenotype. *Cancer Discov.* **6,** 650–663 (2016).

157. Hsu, P. P. *et al.* The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. *Science* **332,** 1317–22 (2011).

158. Rubbi, L. *et al.* Global Phosphoproteomics Reveals Crosstalk Between Bcr-Abl and Negative Feedback Mechanisms Controlling Src Signaling. *Sci. Signal.* **4,** ra18-ra18 (2011).

159. Choudhary, C. *et al.* Mislocalized Activation of Oncogenic RTKs Switches

Downstream Signaling Outcomes. *Mol. Cell* **36,** 326–339 (2009).

160. Klammer, M. *et al.* Phosphosignature Predicts Dasatinib Response in Non-small Cell Lung Cancer. *Mol. Cell. Proteomics* **11,** 651–668 (2012).

161. Alcolea, M. P., Kleiner, O. & Cutillas, P. R. Increased confidence in large-scale phosphoproteomics data by complementary mass spectrometric techniques and matching of phosphopeptide data sets. *J. Proteome Res.* **8,** 3808–3815 (2009).

162. Neilson, K. A. *et al.* Less label, more free: Approaches in label-free quantitative mass spectrometry. *Proteomics* **11,** 535–553 (2011).

163. Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative Proteomic Analysis of Eleven Common Cell Lines Reveals Ubiquitous but Varying Expression of Most Proteins. *Mol. Cell. Proteomics* **11,** M111.014050-M111.014050 (2012).

164. Pierobon, M., Wulfkuhle, J., Liotta, L. & Petricoin, E. Application of molecular technologies for phosphoproteomic analysis of clinical samples. *Oncogene* 1–10 (2014). doi:10.1038/onc.2014.16

165. Hunter, T. Signaling--2000 and beyond. *Cell* **100,** 113–127 (2000).

166. Pawson, T. & Nash, P. Assembly of cell regulatory systems through protein interaction domains. *Science* **300,** 445–452 (2003).

167. Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases A large group of genes in all eukaryotes encode for. *October* **103,** 211–225 (2000).

168. Olsen, J. V. *et al.* Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell* **127,** 635–648 (2006).

169. van ’t Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415,** 530–536 (2002).

170. Chin, K. *et al.* Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* **10,** 529–541 (2006).

171. Wyckoff, J. B., Jones, J. G., Condeelis, J. S. & Segall, J. E. A critical step in metastasis: In vivo analysis of intravasation at the primary tumor. *Cancer Res.* **60,** 2504–2511 (2000).

172. Lehmann, B. D. B. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* **121,** 2750–2767 (2011).

173. Holliday, D. L. & Speirs, V. Choosing the right cell line for breast cancer research. *Breast Cancer Res.* **13,** 215 (2011).

174. Lawrence, R. T. *et al.* The Proteomic Landscape of Triple-Negative Breast Cancer. *Cell Rep.* **11,** 630–644 (2015).

175. Birgersdotter, A., Sandberg, R. & Ernberg, I. Gene expression perturbation in vitro—A growing case for three-dimensional (3D) culture systems. *Semin. Cancer Biol.* **15,** 405–412 (2005).

176. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43,** D512–D520 (2015).

177. Sarbassov, D. D., Guertin, D. a, Ali, S. M. & Sabatini, D. M. Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science* **307,** 1098–1101 (2005).

178. Rikova, K. *et al.* Global Survey of Phosphotyrosine Signaling Identifies Oncogenic Kinases in Lung Cancer. *Cell* **131,** 1190–1203 (2007).

179. Cicenas, J. The potential role of Akt phosphorylation in human cancers. *Int. J. Biol. Markers* **23,** 1–9 (2008).

180. Nakanishi, K., Sakamoto, M., Yamasaki, S., Todo, S. & Hirohashi, S. Akt phosphorylation is a risk factor for early disease recurrence and poor prognosis in hepatocellular carcinoma. *Cancer* **103,** 307–12 (2005).

181. Nam, S. Y. *et al.* Akt/PKB activation in gastric carcinomas correlates with

clinicopathologic variables and prognosis. *APMIS* **111,** 1105–1113 (2003).

182. Yang, X. *et al.* Significance of phosphoinositide 3kinase/AKT pathway alterations in endometrial carcinoma. *Chinese J. Pathol.* **40,** 799–804 (2011).

183. Yang, Y. *et al.* Prognostic value of phospho-Akt in patients with non-small cell lung carcinoma: A meta-analysis. *Int. J. Cancer* **135,** 1417–1424 (2014).

184. Cinti, C. *et al.* Activated Akt as an indicator of prognosis in gastric cancer. *Virchows Arch.* **453,** 449–455 (2008).

185. Hsu, Y. H. *et al.* Definition of PKC-α, CDK6, and MET as therapeutic targets in triple-negative breast cancer. *Cancer Res.* **74,** 4822–4835 (2014).

186. Crockett, D. K., Lin, Z., Vaughn, C. P., Lim, M. S. & Elenitoba-Johnson, K. S. J. Identification of proteins from formalin-fixed paraffin-embedded cells by LC-MS/MS. *Lab. Invest.* **85,** 1405–1415 (2005).

187. Matsuda, K. M., Chung, J. Y. & Hewitt, S. M. Histo-proteomic profiling of formalin-fixed, paraffin-embedded tissue. *Expert Rev Proteomics* **7,** 227–237 (2010).

188. Nirmalan, N. J., Harnden, P., Selby, P. J. & Banks, R. E. Mining the archival formalin-fixed paraffin-embedded tissue proteome: opportunities and challenges. *Mol. Biosyst* **4,** 712–720 (2008).

189. Azimzadeh, O. *et al.* Formalin-fixed paraffin-embedded (FFPE) proteome analysis using gel-free and gel-based proteomics. *J. Proteome Res.* **9,** 4710–4720 (2010).

190. Rimm, D. L. What brown cannot do for you. *Nat. Biotechnol.* **24,** 914–916 (2006).

191. Gavrielides, M. A., Gallas, B. D., Lenz, P., Badano, A. & Hewitt, S. M. Observer variability in the interpretation of HER2/neu immunohistochemical expression with unaided and computer-aided digital microscopy. *Arch. Pathol. Lab. Med.* **135,** 233–242 (2011).

192. Bartholomeusz, C. *et al.* High ERK protein expression levels correlate with shorter survival in triple-negative breast cancer patients. *Oncologist* **17,** 766–74 (2012).

193. Fujii, S. *et al.* MEK-ERK pathway regulates EZH2 overexpression in association with aggressive breast cancer subtypes. *Oncogene* **30,** 4118–28 (2011).

194. Finn, R. S. *et al.* PD 0332991, a selective cyclin D kinase 4/6 inhibitor, preferentially inhibits proliferation of luminal estrogen receptor-positive human breast cancer cell lines in vitro. *Breast Cancer Res.* **11,** R77 (2009).

195. Burstein, M. D. *et al.* Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clin. Cancer Res.* **21,** 1688–98 (2015).